HYDRODYNAMICAL PROCESSES IN ASTROPHYSICAL DISKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ryan Joseph Miranda

December 2017

HYDRODYNAMICAL PROCESSES IN ASTROPHYSICAL DISKS

Ryan Joseph Miranda, Ph.D.

Cornell University 2017

Accretion disks–distributions of gas orbiting a central massive object–are found in a variety of astrophysical settings, such as protoplanetary disks around young stars and disks around supermassive black holes in galactic centers. In this dissertation, I explore several hydrodynamical processes in accretion disks, primarily using numerical hydrodynamics simulations. I study the effect of viscosity on unstable oscillation modes in disks around black holes, showing that it can either suppress or enhance instability. I study the effect of a vortex instability that may arise at the edges of dead zones in protoplanetary disks, showing that it can transport angular momentum through the disk, as well as trap dust, with the latter process resulting in observable morphological features. I investigate the resonant truncation of disks in binaries when they are misaligned with the binary orbital plane, showing that misaligned disks around stars with companions are larger than aligned disks, and that misaligned circumbinary disks generally have smaller cavities than aligned disks. I undertake a comprehensive investigation of the dynamics and evolution of circumbinary disks, with a particular emphasis on the transfer of angular momentum between the disk and binary. I find that circumbinary disks may cause binaries to expand, contrary to the typical expectation that disks tend to harden binaries. Finally, I investigate the migration of a low-mass planets near the inner edge of protoplanetary disk, and show that interactions with the inner edge halt the migration of the planet when its orbital radius is several times larger than the inner disk radius.

BIOGRAPHICAL SKETCH

Ryan Miranda was born in 1989 in Phoenix, AZ, and grew up in the suburb of Cave Creek, AZ. After graduating Cactus Shadows High School in 2007, he attended Arizona State University in Tempe, AZ, where he received a BS in Physics in 2011. He then began pursuing a PhD in Astronomy at Cornell University in Ithaca, NY, working with Professor Dong Lai. After defending his dissertation in August 2017, he will begin a postdoctoral position at the Institute for Advanced Study in Princeton, NJ. Besides astrophysics, Ryan's interest include metal and indie folk music, playing guitar and keyboards, singing karaoke, and mixology.

"I hold the torch ov Heraclitus so I can shake the earth and move the suns"

—Behemoth, "Ov Fire and the Void"

ACKNOWLEDGEMENTS

I thank my advisor, Dong Lai, for guiding the research presented in this dissertation and helping me to develop as a scientist. I acknowledge the contributions of my collaborators and co-authors, Diego Muñoz, Hui Li, Shengtai Li, Héloïse Méheut, and Jiří Horák. I acknowledge the hospitality of Los Alamos National Laboratory during the Summer of 2016, where some of the work presented in this dissertation was carried out. I am thankful for the support of my family, and all of my friends at Cornell, over the past six years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Astrophysical disks are found in a variety of astrophysical contexts, including protoplanetary disks around young stars, disks around compacts objects accreted from binary companions, and disks around supermassive black holes in galactic nuclei. These ubiquitous rotating flows serve as a laboratory for a wide variety of hydrodynamical processes. In this dissertation, I explore several such processes, including hydrodynamical instabilities, angular momentum transport, dust-gas interactions, and binary-disk/planet-disk interactions. Many of these processes have implications for the formation and evolution of planetary systems, but also have applicability to a wider variety of systems, such as coaelescing supermassive black hole binaries. The studies presented in this dissertation have mainly been carried out using two-dimensional numerical hydrodynamics simulations (with the exception of chapter 5, which uses only semi-analytic methods), using several different hydrodynamical codes, including PLUTO (Mignone et al. 2007), FARGO3D (Benítez-Llambay & Masset 2016), and LA-COMPASS (Li et al. 2005). In the remainder of this chapter, I briefly describe the background, context, and main results of each of the remaining chapters.

The transport of angular momentum plays a central role in the evolution of accretion disks. In the classical view of disk evoution, a phenomenological viscosity, often characterized by the viscous α parameter (Shakura & Sunyaev 1973) provides friction between differentially rotating annuli, leading to the outward transfer of angular momentum. As a result, the disk spreads, with the inner regions being accreted by the central star and the outer regions dispersing to

large radii (Lynden-Bell & Pringle 1974). The mechanism behind the effective viscosity was elusive until the discovery of the importance of the magnetorotational instability (MRI; Balbus & Hawley 1991, 1998), which can give rise to turbulence that can effectively transport angular momentum. This mechanism has been a strong candidate for accretion disks which sustain a sufficient level of ionization.

Viscosity also plays a role in the short-term dynamics of disks, impacting the excitation, damping, and propagation of waves. While viscosity often damps waves, it can sometimes excite them due to the "viscous overstability" effect (Kato 1978; Schmit & Tscharnuter 1995; Schmidt et al. 2001). In chapter 2, I numerically investigate the effect of viscosity on a class of oscillation modes in accretion disks around black holes. These "p-modes", related to spiral density waves, are intrinsically unstable due to a general relativistic effect of the rotation profile near the central black hole (Tsang & Lai 2008; Lai & Tsang 2009; Fu & Lai 2011, 2013), and may be the origin of the "quasi-periodic oscillations" (QPOs) in the X-ray light-curves of some accreting black hole binaries (see Remillard & McClintock 2006; Belloni et al. 2012). I find that these oscillations can be either driven by viscosity or somewhat suppressed by it, and that in some cases, viscosity excites a different type of oscillation arising due ot interactions with the inner disk boundary (Tsang & Lai 2009; Fu & Lai 2012). However, the characteristic frequency signature of observed QPOs is not reproduced by these mechanisms.

In protoplanetary disks–the birthplaces of planetary systems–the role of the MRI in transporting angular momentum is uncertain. Shortly after the discovery of the importance of the MRI, it was realized that large regions of proto-

planetary disks are too cold and dense to achieve the necessary levels of ionization required for the MRI to operate (Gammie 1996). This notion was later supported by magnetohydrodynamics simulations including non-ideal effects resulting from the weakly ionized conditions (e.g., Bai & Stone 2013; Lesur et al. 2014; Simon et al. 2015). The cold, inactive regions of protoplanetary disks are known as "dead zones". In the absence of MRI, alternative angular momentum transport mechanisms must operate in order to explain the inferred lifetimes of protoplanetary disks (e.g., Hartmann et al. 1998; Haisch et al. 2001). This is an area of intense, ongoing research (see Armitage 2011; Turner et al. 2014 for reviews).

Among several promising mechanisms for "reviving" the dead zone is the Rossby wave instability (RWI), a global instability of axisymmetric "bumps" in disks (Lovelace et al. 1999; Li et al. 2000, 2001), which can arise at the edges of dead zones. Vortices and density waves produced by the RWI produce fluid stresses which transport angular momentum (Varnière & Tagger 2006; Lyra & Mac Low 2012; Regály et al. 2012). Vortices are also of interest due to their ability to trap dust particles (Barge & Sommeria 1995; Godon & Livio 2000; Méheut et al. 2012), which may facilitate the formation of planetesimals, as well as produce observable morphological features in protoplanetary disks (Lyra & Lin 2013; Zhu & Stone 2014). Chapters 3 and 4 are concerned with the consequences of the RWI excited at dead zone edges.

In chapter 3, I present the results of numerical simulations demonstrating the effectiveness of the RWI at reviving dead zones. This study improves upon existing work by considering a dead zone of finite radial extent (i.e., having both an inner and outer edge), as well as by evolving the disk over a sufficiently long

timescale to verify that a quasi-steady state is reached. The effective viscous α parameter due to the the RWI can be as large as several times 10^{-2}, and it is likely that RWI will revive the dead zone before it accumulates enough mass to become gravitationally unstable.

Several evolved protoplanetary disks, known as transition disks, exhibit strong asymmetries in dust emission at millimeter wavelengths (van der Marel et al. 2013; Casassus et al. 2013; Isella et al. 2013; Pérez et al. 2014). In chapter 4, using the results of hydrodynamical simulations of mutually-coupled dust and gas, I show that these features can be created by RWI vortices at dead zone edges. Most importantly, this study includes the effect of the back-reaction drag of the dust on the gas, which is known to destroy vortices in some circumstances (e.g., Fu et al. 2014; Surville et al. 2016). I find that vortices created at dead zone edges are not destroyed by the dust back-reaction, due to the continuous replenishment of the vortex by the buildup of mass at the dead zone edge. As a result, dust can remain trapped in vortices over timescales comparable to the disk lifetime ($\sim 10^4$ orbits). I create synthetic millimeter wavelength interferometric images demonstrating the observability of these features.

In a variety of astrophysical contexts, a disk-hosting object (star or compact object) may be a member of a binary. Three separate disks may exist in a binary: two circumstellar disks (one around each member of the binary), and a circumbinary disk surrounding the entire binary. The tidal gravitational field of the binary modifies the structure, dynamics, and evolution of both types of disks, by exciting waves which carry angular momentum away from the binary orbit (Goldreich & Tremaine 1979). This effect overwhelms the viscous spreading of the disk and truncates it, in a manner analogous to the opening of a gap

in a protoplanetary disk by a massive planet (e.g., Lin & Papaloizou 1979, 1986). In the case of a circumstellar disk, the outer edge is truncated, while for a circumbinary disk, the inner edge is truncated, resulting in a low-density central cavity surrounding the binary.

The truncation of disks in binaries was first investigated by Artymowicz & Lubow (1994), for the case of disks which are aligned with the orbital plane of the binary. In chapter 5, I extend their analysis to include disks which are misaligned with the binary orbital plane. Such disks are theoretically predicted to exist in young stellar binaries, due to the turbulent formation processes of stars and disks (e.g., Bate, Bonnell & Bromm 2003; McKee & Ostriker 2007; Fielding et al. 2015), and have been detected observationally (e.g., Davis et al. 1994; Winn et al. 2004; Jensen & Akeson 2014). I find that misaligned circumstellar disks have systematically larger outer radii than aligned disks, and are likely to fill their Roche lobes if inclined by more than $45^\circ - 90^\circ$, depending on the binary mass ratio and disk viscosity parameter. Misaligned circumbinary disks generally have smaller inner radii than aligned disk, but the details depend sensitively on the binary and disk parameters.

Circumbinary disks are host to a variety of rich dynamical processes, including periodic variability of the binary accretion rate (e.g., Artymowicz & Lubow 1996; MacFadyen & Milosavljević 2008), disk eccentricity, which has implications for the growth of planets around binaries (Meschiari 2012; Rafikov 2013; Silsbee & Rafikov 2015), and the exchange of angular momentum with the binary through gravitational torques, impacting the orbital evolution of the binary. The angular momentum exchange is thought to play a central role in facilitating the mergers of supermassive black hole binaries binaries (e.g., Begelman

et al. 1980, Armitage & Natarajan 2002; Kelley et al. 2016), which are expected to host circumbinary disks as a consequence of galaxy mergers (e.g., Begelman et al. 1980; Ivanov et al. 1999; Milosavljević & Phinney 2005), providing a possible solution to the so-called "final parsec problem"

In chapter 6, I investigate of the structure and evolution of circumbinary disks using hydrodynamic simulations. This work is restricted to disks which are coplanar with the binary. I improve upon existing numerical simulations of circumbinary accretion (e.g., Artymowicz & Lubow 1996; Günther & Kley 2002; Cuadra et al. 2009; MacFadyen & Milosavljević 2008; Shi et al. 2012; D'Orazio et al. 2013; Farris et al. 2014) by considering both circular and eccentric binaries, and by ensuring that the disk reaches a steady state. I explore the truncation, periodic accretion variability, excitation of coherent eccentric modes in the disk, and angular momentum transfer between the disk and binary. The critical finding of this chapter is that the binary typically receives angular momentum from the disk, implying that the binary orbit expands with time. This is contrary to the expectation that gaseous disks around supermassive black hole binaries may facilitate their mergers.

Planets interact with their natal disks, leading to angular momentum exchange and orbital migration (Goldreich & Tremaine 1979, 1980; Lin & Papaloizou 1979). In chapter 7, I carry out numerical simulations of a low-mass planet undergoing Type I migration (in which the planet does not significantly perturb the disk structure; e.g., Ward 1997) near the inner boundary of a protoplanetary disk. This study is a numerical verification of the theoretical prediction of Tsang (2011), which proposes a modification of classical Type I migration in a disk having a semi-rigid inner edge. The reflection of waves from the inner

edge enhances the positive torque exerted on the planet by material inside its orbit, which can result in the halting of inward migration. Using numerical simulations of planet-disk interaction, I confirm the plausibility of this mechanism, and show it is more effective than predicted by the analytic theory. Planets as massive as about 10 Earth masses can be trapped with a semi-major axis several times larger than the inner disk radius, even for relatively large viscosities. This implies that additional mechanisms besides disk-driven migration may be required to form very short-period planetary systems.

CHAPTER 2

VISCOUS DRIVING OF GLOBAL OSCILLATIONS IN ACCRETION DISKS AROUND BLACK HOLES

2.1 Introduction

High-frequency quasi-periodic oscillations (HFQPOs) in X-ray flux have been observed in a number of black hole X-ray binaries since the 1990s. These low-amplitude ($\sim 1\%$) variabilities occur in the so-called intermediate state or "steep power law" state, and have frequencies of 40 – 450 Hz, comparable to the orbital frequency the innermost stable circular orbit (ISCO) of a $\sim$ 10 $M_\odot$ black hole. This indicates that the physical mechanism behind them is closely related to the dynamics of the innermost regions of black hole accretion disks (for reviews, see Remillard & McClintock 2006 and Belloni et al. 2012). Several classes of models for HFQPOs have been proposed (see Lai & Tsang 2009, Lai et al. 2013 for a review), including the orbital motion of hot spots in the disk (Stella et al. 1999; Schnittman & Bertschinger 2004; Wellons et al. 2014), nonlinear resonances (Abramowicz & Kluźniak 2001; Abramowicz et al. 2007), and oscillations of finite accretion tori (Rezzola et al. 2003; Blaes et al. 2006). A large class of models are based on relativistic diskoseismology, in which the HFQPO frequencies are identified as those of global oscillations modes of the inner accretion disks (see Kato 2001 or Ortega-Rodríguez et al. 2008 for reviews of disk oscillations, Lai et al. 2013 for their connection to HFQPOs). Two possible types of oscillations are inertial-gravity modes (or g-modes) with vertical structure

This chapter is adapted from Miranda, Horák, & Lai (2015).

and inertial-acoustic p-modes with no vertical structure. While the g-modes exhibit a unique self-trapping property, there is some indication that they may be destroyed by subthermal magnetic fields (Fu & Lai 2009), comparable to those produced by saturation of the magnetorotational instability, or be destroyed by turbulence (Reynolds & Miller 2009). Therefore it is unclear whether or not g-modes can exist in real disks. However, p-modes are only weakly affected by magnetic fields (Fu & Lai 2011), and their lack of vertical structure makes them insensitive to turbulence.

The types of oscillations which are possible in the disk depend on the properties of its inner edge. Either the gas freely flows into the central black hole, or there exists an inner edge which is impermeable to radial motions, for example due to the presence of a magnetosphere (e.g. Bisnovatyi-Kogan & Ruzmaikin 1974, 1976; Igumenshchev et al. 2003; Rothstein & Lovelace 2008; McKinney et al. 2012). In the latter case, there exist "trapped" p-modes which can propagate between the reflective inner boundary and their inner Lindblad resonances. A trapped p-mode can be overstable if it can draw negative energy from the background flow through the corotation resonance. This is possible if the radial derivative of vortensity (vorticity $|\nabla\times\mathbf{v}|$ divided by surface density) is positive at the corotation radius, which, in the absence of sharp changes in surface density, is not possible in Newtonian disks, but can be achieved with general relativistic (GR) effects (see Tsang & Lai 2008; Lai & Tsang 2009; Horák & Lai 2013). This "corotational instability" has been demonstrated in 2D simulations of inviscid disks (Fu & Lai 2013).

One of the goals of this paper is to examine the effect of viscosity on trapped p-modes. In particular, we are interested in whether viscosity suppresses or en-

hances their growth. The latter may occur as the result of viscous overstability. We model the disk turbulent viscosity by the standard α-ansatz, with different equations of state. The classic theory of viscous overstability describes an axisymmetric instability in which viscosity injects energy drawn from the disk shear into growing oscillations (Kato 1978; Schmit & Tscharnuter 1995; Schmidt et al. 2001). This effect requires viscous forces to act in phase with oscillations, which is achieved if the dynamic viscosity η scales sufficiently steeply with surface density Σ, i.e., if the parameter $A = \mathrm{d}\ln\eta/\mathrm{d}\ln\Sigma$ is larger than unity. Viscous overstability has been demonstrated in both hydrodynamic (Latter & Ogilvie 2010) and N-body simulations (Rein & Latter 2013). Extensions of the original axisymmetric theory show that a similar effect can drive non-axisymmetric oscillations (Papaloizou & Lin 1988; Lyubarskij et al. 1994). Previous analytical calculations, based on local analysis (see Section 2.2), indicate that viscosity can excite disk p-modes due to viscous forces acting in phase with the oscillation (e.g. Ortega-Rodríguez & Wagoner 2000; Kato 2001). However, local analysis cannot determine the damping or growth rate of global modes.

In the absence of a reflective boundary, the inner edge of the disk is free-flowing, with gas interior to the ISCO plunging into the central black hole, forming a transonic radial flow. Kato et al. (1988a) found that the sonic point of such flow is unstable to axisymmetric perturbations for sufficiently large viscosities, a mechanism which is distinct from the standard viscous overstability. One-dimensional simulations of transonic disks have shown that viscosity can drive global oscillations at the maximum epicyclic frequency κ_{max} (Milsom & Taam 1996; Mao et al. 2009). Two-dimensional and three-dimensional simulations have shown global oscillations at the same frequency (O'Neill et al. 2009), as well as at multiples of innermost stable orbital frequency Ω_{ISCO} (Chan 2009).

Another goal of this paper is to study the conditions for producing axisymmetric and various non-axisymmetric modes in transonic disks.

The outline of this paper is as follows. In Section 2.2, we review the theory of viscous overstability, deriving the formula for the growth rate of local non-axisymmetric perturbations due to viscosity. We also provide estimates of the global growth rates of trapped p-modes under a pseudo-Newtonian potential. The setup for our numerical experiments is described in Section 2.3. In Section 2.4, we present the results of simulations of overstable oscillations in a transonic disk with a free inner boundary. We classify several types of global oscillation modes produced in this flow, as well their dependence on the sound speed and viscosity parameter of the disk. In Section 2.5 we examine trapped p-modes in a disk with a reflecting boundary, first reproducing semi-analytic results for an inviscid disk, then investigating the effect of viscosity as described by the theory of viscous overstability. As part of this discussion (Section 2.5.3), we present an unexpected result on overstable modes driven purely by viscosity in the absence of GR effects, whose frequencies are higher than trapped p-modes. We summarize our results and discuss their implications in Section 2.6.

2.2 Theory of Viscous Instability

The role of viscosity for the stability of thin accretion disks has been studied extensively for several decades. As shown by Kato (1978) in his pioneering work, viscosity can affect the disk stability through both thermal (affecting the energy balance) and dynamical processes (by changing the angular momentum balance). The former process is analogous to so called 'ϵ-mechanism' in stellar

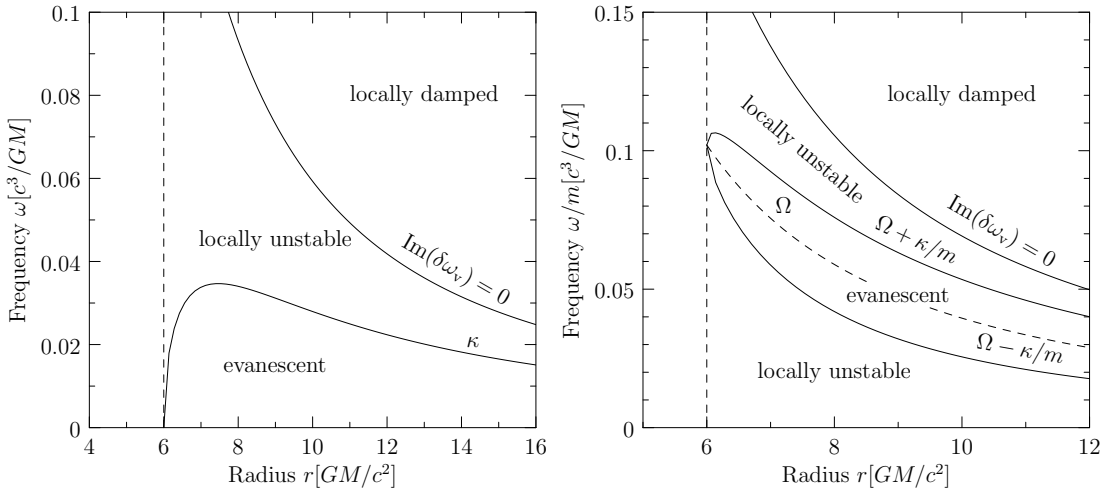

Figure 2.1: Regions of local viscous stability and instability in the propagation diagram of axisymmetric (left) and non-axisymmetric p-waves. The 'evanescent' regions correspond to the cases where $D > 0$, the 'locally damped' and 'locally unstable' regions correspond to positive or negative ω_{v}, respectively.

pulsations. In that case, the instability arises as a consequence of an additional increase in the pressure restoring force, due to viscous heat generation in the compressed phase of the oscillations. In the later case, the instability comes from the mechanical work that is done by the azimuthal component of the viscous force on the fluid elements during the oscillations. To explore this mechanism, it is sufficient to consider barotropic fluid, which we adopt in our work.

We consider a thin, two-dimensional disk described in polar coordinates (r, ϕ) by velocity $\boldsymbol{u} = (u_r, u_\phi)$, surface density Σ and height integrated pressure P which obey the continuity and Navier-Stokes equations

$$\frac{\partial \Sigma}{\partial t} + \boldsymbol{\nabla} \cdot (\Sigma \boldsymbol{u}) = 0, \tag{2.1}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \boldsymbol{\nabla})\, \boldsymbol{u} = -\frac{1}{\Sigma} \boldsymbol{\nabla} P - \boldsymbol{\nabla} \Phi + \frac{1}{\Sigma} \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}, \tag{2.2}$$

where $\boldsymbol{\sigma}$ is the viscous stress tensor,

$$\boldsymbol{\sigma} = \eta \left[\boldsymbol{\nabla} \boldsymbol{v} + \boldsymbol{v} \boldsymbol{\nabla} - \frac{2}{3} (\boldsymbol{\nabla} \cdot \boldsymbol{v})\, \mathbf{I} \right], \tag{2.3}$$

$\Phi = \Phi(r)$ is the gravitational potential and η is the height integrated dynamic viscosity coefficient that describes momentum transport due to turbulent motion of the fluid. In this section, it is sufficient to assume that both P and η are functions of the surface density only, $P = P(\Sigma)$ and $\eta = \eta(\Sigma)$. The particular forms of these functions will be specified later, when the numerical simulations are carried out.

2.2.1 Subsonic Part of Disk and Trapped Modes

In absence of the viscosity ($\eta = 0$), the equations (2.1) and (2.2) admit a stationary axisymmetric solution, that describes a purely rotating flow with the squared angular velocity

$$\Omega^2 = \Omega_{\mathrm{K}}^2 + \frac{c_{\mathrm{s}}^2}{r^2}\frac{\mathrm{d}\ln\Sigma}{\mathrm{d}\ln r}, \quad \Omega_{\mathrm{K}}^2 = \frac{1}{r}\frac{\mathrm{d}\Phi}{\mathrm{d}r}, \tag{2.4}$$

where Ω_{K} is the Keplerian angular frequency and $c_{\mathrm{s}} = (\mathrm{d}P/\mathrm{d}\Sigma)^{1/2}$ is the sound speed. Its contribution to the rotational velocity is negligible in thin disks giving nearly Keplerian rotation, $\Omega \approx \Omega_{\mathrm{K}}$.

The perturbations to this equilibrium, whose time and azimuthal dependence is of the form of $\exp(\mathrm{i}m\phi - \mathrm{i}\omega t)$, obey the wave equation (e.g. Lai & Tsang 2009),

$$\begin{aligned}\hat{L}^0\delta h = \Bigg[&\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \left(\frac{\mathrm{d}}{\mathrm{d}r}\ln\frac{D}{r\Sigma}\right)\frac{\mathrm{d}}{\mathrm{d}r} \\ &+ \frac{2m\Omega}{r\tilde{\omega}}\left(\frac{\mathrm{d}}{\mathrm{d}r}\ln\frac{D}{\Omega\Sigma}\right) - \frac{m^2}{r^2} - \frac{D}{c_{\mathrm{s}}^2}\Bigg]\delta h = 0\end{aligned} \tag{2.5}$$

for the enthalpy perturbation $\delta h = \delta P/\Sigma$. Here ω and m are the angular frequency and azimuthal wavenumber of the perturbation, $\tilde{\omega} = \omega - m\Omega$, $D = \kappa^2 - \tilde{\omega}^2$ and

$$\kappa^2 = r\frac{\mathrm{d}\Omega^2}{\mathrm{d}r} + 4\Omega^2 \tag{2.6}$$

is the squared radial epicyclic frequency. The perturbation of the flow velocity $\delta\boldsymbol{u}$ are given as

$$\delta u_r = \frac{\mathrm{i}}{D}\left[\tilde{\omega}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{2m\Omega}{r}\right]\delta h, \quad \delta u_\phi = \frac{1}{D}\left[\frac{\kappa^2}{2\Omega}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{m\tilde{\omega}}{r}\right]\delta h. \tag{2.7}$$

The operator $\hat{L}^0$ is singular when $D \to 0$ or $\tilde{\omega} \to 0$. The first corresponds to the Lindblad resonances and the second to the corotation resonance. While the Lindblad resonances are not real singularities and represent turning points of the waves, the waves may be absorbed at the corotation resonance, which may lead to the instability or damping of the oscillations of the disk.

Introducing the viscosity changes both the equlibrium state and the dynamics of the oscillations. The angular momentum transport causes a slow radial inflow of the matter towards the central black hole in the stationary case. The radial velocity u_r of this inflow can be found by expanding the azimuthal component of equation (2.2) up to the first order in η,

$$u_r = \frac{2\Omega}{r^2\kappa^2\Sigma}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^3\eta\frac{\mathrm{d}\Omega}{\mathrm{d}r}\right). \tag{2.8}$$

This expansion however breaks down close to ISCO, where $\kappa \to 0$ and the flow becomes transonic. The stability of this region of the disk is reviewed later in this section and numerically examined in section 4.

In addition to a small change in the mean flow, the viscosity affects the oscillations through the perturbation of the viscous force $\delta(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma})$ that may do a positive or negative work on the waves. Up to the first order in viscosity, the perturbations of the disk are governed by the equation (see Appendix A.1)

$$\left(\hat{L}^0 + L^1_{\mathrm{v}} + L^1_{\mathrm{in}}\right)\delta h = 0, \tag{2.9}$$

where $\hat{L}^0$ is the differential operator of the inviscid problem defined in equation (2.5) and $\hat{L}^1_{\mathrm{in}}$ and $\hat{L}^1_{\mathrm{v}}$ are its first-order corrections due to the radial inflow in the

stationary case and the action of the viscous force:

$$\hat{L}^1_{\rm in} = {\rm i}u_r \frac{\tilde{\omega}}{D}\left[\left(1+\frac{\kappa^2}{\tilde{\omega}^2}\right)\frac{{\rm d}^3}{{\rm d}r^3} - \frac{D^2}{\tilde{\omega}^2 c_{\rm s}^2}\frac{\rm d}{{\rm d}r}\right], \tag{2.10}$$

$$\hat{L}^1_{\rm v} = -{\rm i}\nu \frac{\tilde{\omega}}{D}\left[\left(\frac{4}{3}+\frac{\kappa^2}{\tilde{\omega}^2}\right)\frac{{\rm d}^4}{{\rm d}r^4} - 2qA\frac{D}{\tilde{\omega}^2}\frac{\Omega}{c_{\rm s}^2}\frac{{\rm d}^2}{{\rm d}r^2}\right], \tag{2.11}$$

where $\nu = \eta/\Sigma$ is the kinematic viscosity.

By substituting the local WKBJ ansatz, $\delta h \propto \exp({\rm i}kr)$ we may recover the formula of Kato (1978) for the local growth rate of the oscillations due to the action of the viscous force,

$$\delta\omega_{\rm v}(r) = -{\rm i}\nu k^2\left[\frac{2}{3}+\frac{\Omega^2}{\tilde{\omega}^2}\left(\frac{\kappa^2}{2\Omega^2}-qA\right)\right], \tag{2.12}$$

where $q = -{\rm d}\ln\Omega/{\rm d}\ln r$ and $A = {\rm d}\ln\eta/{\rm d}\ln\Sigma$. Because $q > 0$, the last term reduces the damping and in principle it can lead to an instability. A necessary condition for mode growth [${\rm Im}\,(\delta\omega_{\rm v}) > 0$] is positive A, i.e. the shear viscosity coefficient η increases with increasing density Σ on a timescale shorter than the oscillation period ($\sim 1/\Omega$). For a given oscillation frequency, we may separate the regions of local instability or damping of the p-waves in the propagation diagram. This is done in Figure 2.1.

A similar analysis in the case of the operator $\hat{L}_{\rm in}$ gives the correction due to the radial inflow in the disk,

$$\delta\omega_{\rm in}(r) = -ku_r. \tag{2.13}$$

The meaning of this result is straightforward; it is the Doppler shift between the stationary observer at a fixed radius and an observer comoving with the fluid. Hence, the inflow only slightly changes the real part of the frequency of the oscillations without affecting their stability. The total frequency change is the sum of the two contributions, $\delta\omega(r) = \delta\omega_{\rm v}(r) + \delta\omega_{\rm in}(r)$.

Similarly, we may derive a *global* change of the eigenfrequencies of the *p*-modes trapped between two boundaries. A simple calculation presented in Appendix A.2 leads to the expressions

$$\delta\omega_{\mathrm{v,global}} = \frac{\int \delta\omega_{\mathrm{v}}(r)w(r)\mathrm{d}r}{\int w(r)\mathrm{d}r}, \quad \delta\omega_{\mathrm{in,global}} = \frac{\int \delta\omega_{\mathrm{in}}(r)w(r)\mathrm{d}r}{\int w(r)\mathrm{d}r}, \tag{2.14}$$

where

$$w(r) = \frac{1}{c_{\mathrm{s}}}\frac{\tilde{\omega}}{\sqrt{-D}}. \tag{2.15}$$

Hence, the effects of viscosity and radial inflow are just the average of the local rates of equations (2.12) and (2.13) with the weight function $w(r)$.

2.2.2 Transonic flow

The studies of the stability of the transonic regions of accretion disks were initiated by the work of Kato et al (1988a). By perturbing equations (2.1) and (2.2) around a transonic stationary solution describing an isothermal accretion disk, the authors identified two types of unstable axisymmetric perturbations. The first type is essentially a generalization of the inertial-acoustic *p*-waves discussed in the previous subsection to the case of a nonzero radial flow velocity. These modes are trapped between the disk inner edge (corresponding to the sonic radius) and the radius of the inner Lindblad resonance. Consequently, the frequency of these waves is always smaller than κ_{max}, we note however, that this limit is valid only in the case of the axisymmetric perturbations. According to Kato et al. (1988a), as the sonic radius acts only as a partial reflector for the waves, the instability appears only for a sufficiently high viscosity when the viscous driving overcomes the leakage of wave energy through the sonic radius.

The other type of instability appears only in the case of a transonic flow

and represents a standing wave pattern localized around the sonic radius that grows exponentially with time. Kato et al. (1988a) assumed so-called conventional or 'αp'-type viscosity, in which the $r\phi$-component of the viscous stress tensor is directly proportional to the pressure, $\sigma_{r\phi} = \alpha p$ (contrary to our 'diffusive' prescription for the viscosity, that relates pressure to the shear viscosity coefficient η). They found that the instability occurs only for high enough values of α, when $\alpha \geq u'_c/\Omega_c$, with u'_c and Ω_c being a radial velocity gradient and orbital frequency of the flow at the sonic radius. They also noted that the same condition is satisfied when the flow changes from being subsonic to supersonic by passing through the nodal critical point. They speculated that the appearance of this second type of instability is directly related to the topology of the flow at the sonic radius. This idea was supported in subsequent studies: Kato et al. (1988b) relaxed the assumption of isothermal flow by including the energy balance between viscous heating and radiative cooling and found that the instability criterion coincides with the one for the nodal topology of the sonic point. Later on, Kato et al (1993) examined a stability of the isothermal accretion flows with the diffusive form of the viscosity similar to our work. In that case the sonic point is always of the saddle type and consequently the authors found that it is stable against this type of perturbation.

Based on these results we expect that possible instabilities of our flow arise only due to propagating acoustic waves mentioned in the beginning of this subsection. In the following sections, we will examine numerically the conditions for their growth as well as the spatial structure of the unstable modes.

2.3 Numerical Setup

As in the previous section, we consider a thin disk described by velocity $\mathbf{u} = \left(u_r, u_\phi\right)$, surface density Σ and pressure P which obey the Navier-Stokes equations. It is subject to the "pseudo-Newtonian" (Paczyński & Wiita 1980) gravitational potential

$$\Phi = -\frac{GM}{r - r_\mathrm{s}}, \tag{2.16}$$

where M is the mass of the central black hole and $r_\mathrm{s} = 2GM/c^2$ is its Schwarzschild radius. This potential mimics GR effects, having a Keplerian orbital frequency and radial epicyclic frequency given by

$$\Omega_\mathrm{K} = \sqrt{\frac{GM}{r^3}}\left(\frac{r}{r - r_\mathrm{s}}\right), \quad \kappa = \Omega_\mathrm{K}\sqrt{\frac{r - 3r_\mathrm{s}}{r - r_\mathrm{s}}}. \tag{2.17}$$

There is an innermost stable circular orbit (ISCO) defined by $\kappa^2\left(r_\mathrm{ISCO}\right) = 0$, and a radius at which the epicyclic frequency peaks, $\kappa\left(r_\mathrm{max}\right) = \kappa_\mathrm{max} = \left(9 - 5\sqrt{3}\right)\Omega_\mathrm{ISCO}$, which are located at $r_\mathrm{ISCO} = 3r_\mathrm{s}$ and $r_\mathrm{max} = \left(2 + \sqrt{3}\right)r_\mathrm{s}$, respectively.

We adopt numerical units such that

$$r_\mathrm{ISCO} = \Omega_\mathrm{ISCO} = 1, \tag{2.18}$$

and define an "orbit" as one orbital period at r_ISCO (equal to 2π in numerical units). We use a polytropic equation of state $P = K\Sigma^\Gamma$, which has the corresponding sound speed $c_\mathrm{s} = \sqrt{\Gamma P/\Sigma}$. We define the parameter $c_{\mathrm{s}0} = c_\mathrm{s}\left(r_\mathrm{ISCO}\right)$, whose value we will refer to rather than that of the constant K, to which it is directly related. We prescribe the kinematic viscosity of the fluid as

$$\nu = \alpha c_\mathrm{s} H, \tag{2.19}$$

where $H = c_\mathrm{s}/\Omega_\mathrm{K}$ is the disk scale height (Shakura & Sunyaev 1973). Note that ν is proportional to c_s^2, so that $\nu\left(\Sigma\right) \propto \Sigma^{\Gamma-1}$ or $\eta\left(\Sigma\right) \propto \Sigma^\Gamma$, giving $A = \mathrm{d}\ln\eta/\mathrm{d}\ln\Sigma = \Gamma$. Therefore each simulation is defined by three parameters: $c_{\mathrm{s}0}$, Γ and α.

The outer boundary is always located at $r = 4$, at which we fix Σ and u_ϕ at their initial values and impose $\partial u_r/\partial r = 0$. Adjacent to the outer boundary, we implement a wave damping zone of width 1/3 (i.e., between $r = 11/3$ and $r = 4$) in which a damping force (per unit mass), given by

$$f_{\rm damp} = -\frac{\mathbf{u} - \mathbf{u}_0}{\tau} R(r) \tag{2.20}$$

is applied (de Val-Borro et al. 2006). Here $\mathbf{u}_0 = (r\Omega_0, 0)$, τ is a damping timescale equal to the orbital period at the outer boundary, and $R(r)$ is a parabolic function, which is equal to unity at $r = 4$ and reduces to zero at $r = 11/3$. The damping zone minimizes reflection of waves from the outer boundary and mimics an outgoing wave boundary condition used in linear analysis of disk modes (e.g., Lai & Tsang 2009). We simulate two distinct physical setups which correspond to two different inner boundary locations and their corresponding boundary conditions, choosing either a free boundary at $r = 2/3$ (Section 2.4) or a reflecting boundary at $r = 1$ (Section 2.5). The details and significance of these boundary conditions are discussed in more detail in their corresponding sections. The initial conditions of the simulations always have $u_r = 0$ everywhere, but the initial surface density profile $\Sigma(r)$ varies in the different sections of this paper. The orbital velocity u_ϕ is always chosen to be initially in centrifugal balance (equation 2.4) given the surface density profile.

The Navier-Stokes equations are solved using the PLUTO code (Mignone et al. 2007) with third-order Runge-Kutta time stepping, parabolic reconstruction and a Roe solver on static polar grid with uniform spacing in both directions. The standard resolution of the simulations is $N_r \times N_\phi = 1024 \times 256$, so that the radial direction has a spatial resolution of at least 300 per unit r (this varies slightly with the location of the inner boundary), and modes with $m \leq 4$ are captured with at least 64 zones per azimuthal wavelength. The sensitivity of

Label	c_{s0}	α	r_c	u_c'/Ω_c	κ_{max}	$m\Omega_{ISCO}$	$m\Omega_{ISCO} \pm \kappa_{max}$
(a)	0.01	0.05	0.99	0.150	No	No	No
(b)	0.01	0.10	1.02	0.120	No	4	No
(c)	0.01	0.25	1.12	0.087	Linear	2	Yes
(d)	0.01	0.50	1.30	0.082	Nonlinear	1	Yes
(e)	0.02	0.10	1.01	0.170	No	No	No
(f)	0.02	0.25	1.17	0.109	No	1	Yes
(g)	0.02	0.50	1.69	0.116	Nonlinear	No	No
(h)	0.05	0.25	1.12	0.197	No	No	No
(i)	0.05	0.50	1.57	0.446	Nonlinear	No	No

Table 2.1: Summary of transonic disk simulations. The first three columns give the name of the run and the value of parameters c_{s0} and α. The next two columns give the numerically measured location of the sonic point r_c and its dimensionless stability quantity u_c'/Ω_c. The last three columns indicate whether or not overstable oscillations of the three types described in Section 2.4.1 are present. For the κ_{max} type, we also indicate whether or not they are linear or nonlinear, and if the $m\Omega_{ISCO}$ type are present, we indicate which azimuthal number m is dominant in the power spectrum.

our numerical results to the chosen resolution is addressed seperately for each major results section (Sections 2.4 and 2.5).

2.4 Overstable Global Oscillations in Transonic Disks

For our transonic disk simulations, the inner boundary is placed at $r_{in} = 2/3$, at which an outflow boundary condition given by

$$\frac{\partial \Sigma}{\partial r} = \frac{\partial \mathbf{u}}{\partial r} = 0 \tag{2.21}$$

is imposed. We choose as our initial condition

$$\Sigma \propto \frac{r^{-\frac{1}{2}}}{r - r_s}\left[1 - \left(\frac{r}{r_{in}}\right)^{-\frac{1}{2}}\right], \tag{2.22}$$

which is an approximate steady-state surface density profile consistent with constant $\dot{M}$ and vanishing torque at r_{in}, but it is not a solution for the transonic

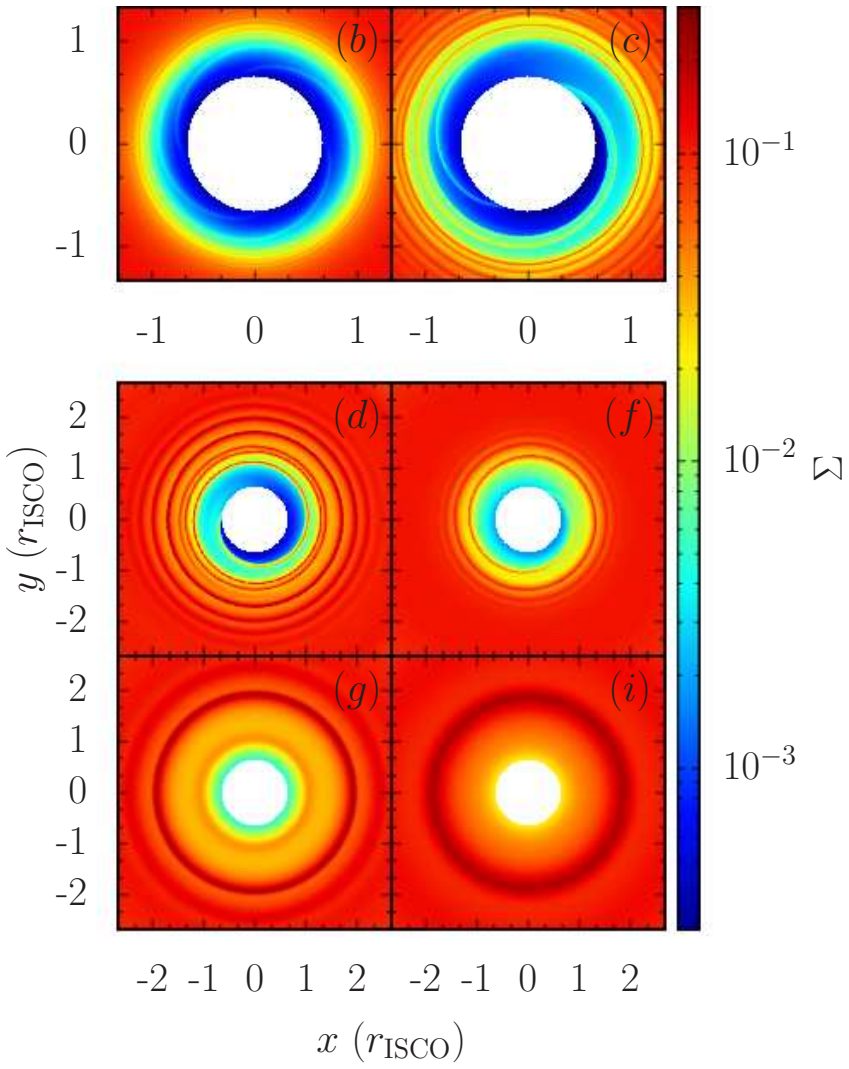

Figure 2.2: Surface density at $t = 100$ for runs with significant time variability. The parameters corresponding to the alphabetical labels are given in Table 2.1. Various features can be seen, including four-armed, two-armed and one-armed spirals, as well as axisymmetric rings. The top two panels are zoomed in relative to the other panels to show finer detail of the non-axisymmetric structure close to the inner edge.

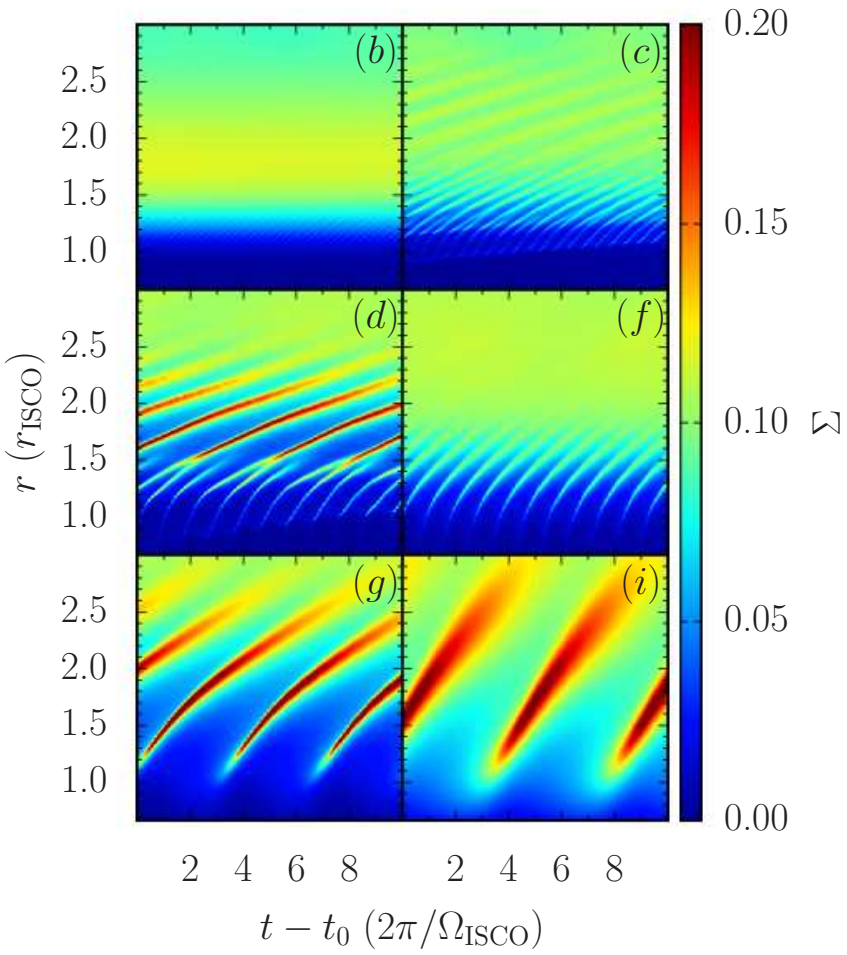

Figure 2.3: Space-time diagrams of surface density Σ at $\phi = 0$, demonstrating the structure of the spatial variabilities and temporal variabilities of the oscillations. Time is shown relative to the reference time $t_0 = 90$. We can see that waves with frequencies close to multiple of $\Omega_{\rm ISCO}$ (if they are present) do not propagate at radii much larger than $r_{\rm ISCO}$, beyond which only axisymmetric oscillations with frequency approximately $\kappa_{\rm max}$ propagate.

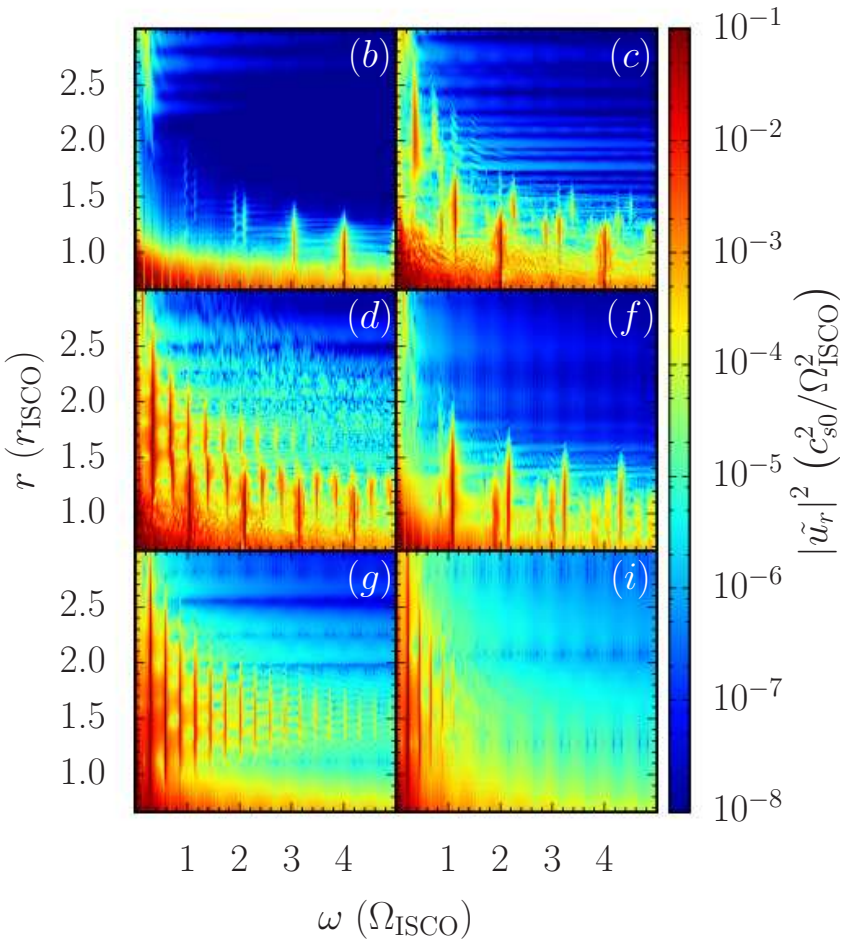

Figure 2.4: The power spectrum $|\tilde{u}_r|^2$ over a range of disk radii. Global modes with power in discrete frequencies which are coherent across a range of r are present in all cases. They can be seperated into modes with frequencies close to multiples of $\Omega_{\rm ISCO}$ (and their nonlinear splittings) with power concentrated near $r_{\rm ISCO}$, and those with frequencies close to multiples of $\kappa_{\rm max}$ with power concentrated at larger radii. These correspond respectively to the spirals and axisymmetric rings seen in Figure 2.2.

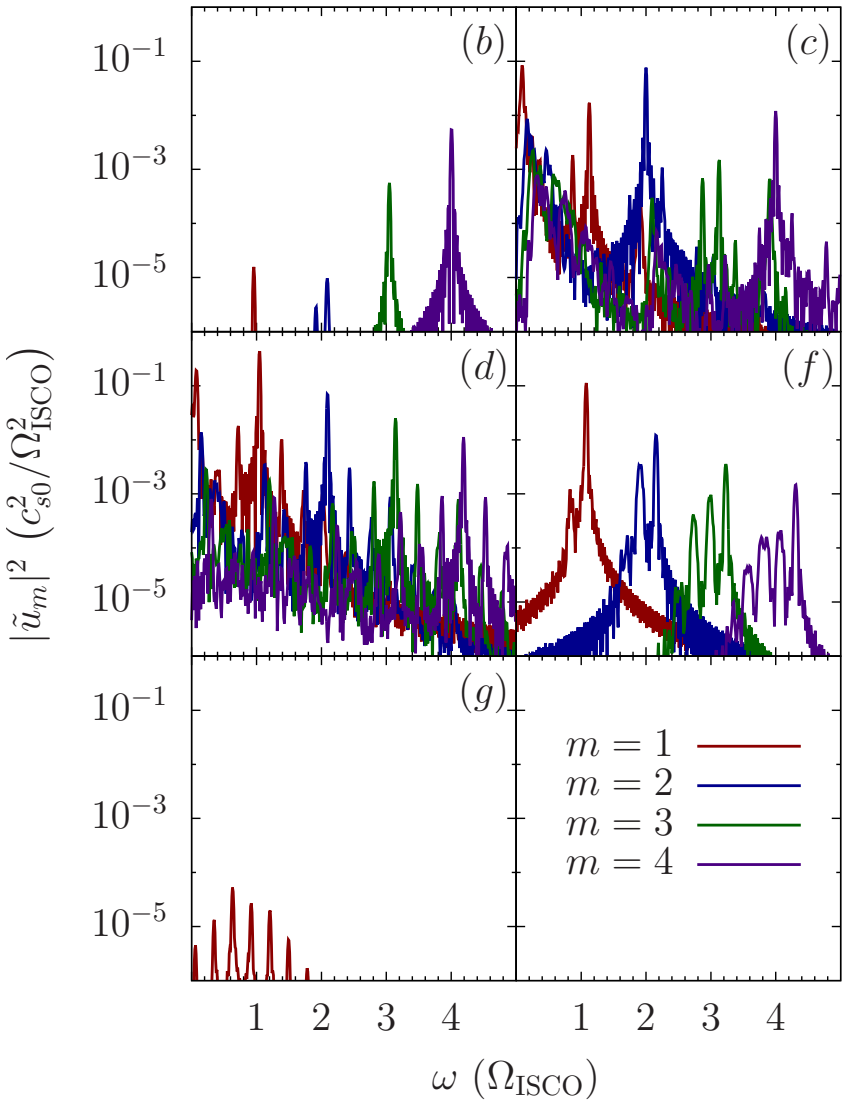

Figure 2.5: Power spectrum delineated by azimuthal number m, given by the quantity $|\tilde{u}_m|^2$ defined in eq. (2.24).

flow region interior to $r_{\rm ISCO}$. From these initial conditions, the disk is evolved for 100 orbits. We wait for it to settle into a quasi-steady state before analyzing its variabilities. The parameters for various runs and their corresponding key results are summarized in Table 2.1. Run (*a*), which has the longest viscous timescale, is the slowest to reach a quasi-steady state, taking $\sim$ 60 orbits (however, this is much less than the viscous timescale). Therefore we perform all analysis in the last 30 orbits of each run, at which point all have reached a quasi-steady state. Since we ignore the initial transient phase, we do not attempt to characterize the growth process or linear behavior of any observed oscillations. Instead, we focus on the properties of unstable oscillations which have saturated at some (often large) amplitudes and which may have become significantly nonlinear.

We use the following formalism to analyze the variabilities of the quasi-steady, nonlinear phase of the simulations. We define the power spectrum $|\tilde{u}|^2$, where

$$\tilde{u}\left(r,\omega\right)=\frac{1}{2\pi}\int_{t_1}^{t_2}u_r\left(r,0,t\right)e^{-i\omega t}\mathrm{d}t \tag{2.23}$$

is the Fourier transformed radial velocity and $(t_1,t_2)=(70,100)$. The power spectrum gives a measure of the strength of the oscillations, along with their frequencies and locations in the disk where they are most visible. We determine the azimuthal number m associated with each frequency by decomposing the radial velocity into components proportional to $\exp\left(im\phi-i\omega t\right)$. Since, as we will demonstrate, the oscillations are global (their frequencies are coherent across a range of radii), we average this amplitude over all r. The resulting quantity,

$$\begin{aligned}\tilde{u}_m\left(\omega\right)=&\frac{1}{4\pi^2\left(r_{\rm out}-r_{\rm in}\right)}\\&\times\int_{r_{\rm in}}^{r_{\rm out}}\int_{t_1}^{t_2}\int_0^{2\pi}u_r\left(r,\phi,t\right)e^{-\left(im\phi-i\omega t\right)}\mathrm{d}\phi\mathrm{d}t\mathrm{d}r,\end{aligned} \tag{2.24}$$

specifies the global power spectrum delineated by m. The spatial and temporal variability and propagation properties of the oscillations are examined by plotting the surface density Σ at a fixed azimuthal angle ($\phi = 0$) as a function of radius r and time t.

For this section we restrict the equation of state to only the isothermal case, $\Gamma = 1$, so that each run is characterized only by the parameters c_{s0} and α. The parameters of the nine numerical runs (see Table 2.1) were chosen because they lead to a variety of behaviors, producing overstable oscillations of several types (to be described below) which are present or absent in various combinations. For each sound speed c_{s0}, there is a threshold value of the viscosity parameter α below which no overstable oscillations are observed within the duration of the run (100 orbits). Three of our runs, (*a*), (*e*) and (*h*), fall into this regime, leaving six runs which show significant variabilities. The subsequent discussion focuses only on these six interesting cases.

Figure 2.2 shows snapshots of the disk surface density Σ at $t = 100$ for the runs that exhibit overstable oscillations. Figure 2.3 shows Σ at $\phi = 0$ as function of time and radius. Together they illustrate the spatial structure and wave propagation properties of the modes. In runs (*b*), (*c*), (*d*) and (*i*), non-axisymmetric waves with frequencies approximately multiples of $\Omega_{\rm ISCO}$ propagate at small radii ($r \lesssim 1.2$ for $c_{s0} = 0.01$, $r \lesssim 1.7$ for $c_{s0} = 0.02$). In runs (*c*), (*d*), (*g*) and (*i*), axisymmetric waves with lower frequencies propagate at larger radii. In cases in which both types of waves are present, there is a merging or winding-up of non-axisymmetric waves into axisymmetric ones at intermediate radii.

Figure 2.4 depicts the power spectra, for a range of radial locations in the disk, of the runs which exhibit overstable oscillations. In all cases, power is

sharply concentrated in discrete frequencies and present across a large range of r, manifesting as narrow vertical strips in Figure 2.4. We therefore interpret these as frequencies of coherent global modes, rather than local oscillations whose properties vary with r. Since they are global, with frequencies independent of location, we can justify the radial integration in equation (2.24), which allows us to analyze their relative strengths globally, rather than locally. The m-delineated power spectra are shown in Figure 2.5. Note that the absolute scale of the quantity $|\tilde{u}_m|^2$ does not have a direct physical meaning, as it depends on the radial integration range (which we choose as the entire computational domain, whose size is arbitrary). However the relative amplitudes, both between different values of m and across multiple panels in Figure 2.5, give a meaningful comparison of the power in the various modes. From these we see that run (b), (c), (d) and (f) are dominated by $m = 4$, $m = 2$, $m = 1$ and $m = 1$ modes, respectively. Runs (g) and (i) have very little power in non-axisymmetric modes, with the majority of their power in axisymmetric ($m = 0$) modes, whose corresponding $|\tilde{u}_m|$ is not shown.

2.4.1 Classification of Oscillations

We classify the observed oscillation modes into three types. First, there is a "$\kappa_{\max}$" type, which is axisymmetric ($m = 0$) and has a frequency approximately equal to (for small c_{s0}) or somewhat smaller than (for large c_{s0}) the maximum epicyclic frequency, $\kappa_{\max} = 0.34$. Under various conditions, this mode is either linear or nonlinear. In the former case, its wave function is smooth and has a small amplitude, and only the fundamental frequency is present in the power spectrum. In the latter case, the wave function is sharp and large in amplitude,

and many overtones (integer multiples of $\kappa_{\max}$) are present in the power spectrum. This behavior is analogous to that of a nonlinear oscillator with natural frequency $\kappa_{\max}$. As the amplitude of the oscillation becomes large, nonlinear effects introduce resonances at multiples of the natural frequency, and oscillations at these frequencies can also be driven to large amplitudes. The result is the excitation of a series of harmonics of the frequency $\kappa_{\max}$. In the most extreme examples of this nonlinearity [e.g. runs (*d*) and (*g*)], the power density (see Figure 2.4) in the first three to four overtones is comparable to (less than, but within an order of magnitude of) that of the fundamental frequency, and up to ten overtone frequencies can be identified [run (*g*)].

The second type of mode is the "$m\Omega_{\rm ISCO}$" type, which is non-axisymmetric with azimuthal number $m > 0$ and frequency $\omega \approx m\Omega_{\rm ISCO}$. These always appear in a series of successive m, and therefore represent a harmonic series of the fundamental frequency (similar to the nonlinear effect seen in the $\kappa_{\max}$ modes), which is simply $\Omega_{\rm ISCO}$. However, the m with the largest power is not always 1, we find cases in which it is 2 or 4. This manifests as a sharp, nonlinear m-armed spiral in the disk.

The third type of mode is the "$m\Omega_{\rm ISCO} \pm \kappa_{\max}$" type, which has azimuthal number $m > 0$ and frequency approximately $m\Omega_{\rm ISCO} \pm \epsilon$, where ϵ is close to, but sometimes less than, $\kappa_{\max}$. These are prominently present for moderate values of α when $c_{\rm s0}$ is small [runs (*c*) and (*d*)], moderately present for a similar α at a larger $c_{\rm s0}$ [run (*f*)] and entirely absent for large α or large $c_{\rm s0}$ [runs (*g*) and (*i*)]. We suggest that these modes are a result of a splitting of the $m\Omega_{\rm ISCO}$ modes due to a nonlinear coupling with the $\kappa_{\max}$ modes, as these are the most fundamental modes associated with the two "special" locations in the disk, $r_{\rm ISCO}$ and $r_{\max}$.

2.4.2 Dependence on Parameters

From our runs, we can extrapolate the dependence on the parameters c_{s0} and α in determining power spectrum of overstable modes. For a given sound speed c_{s0}, we imagine slowly increasing the value of α, starting from a very small value. A sufficiently small α produces no overstable oscillations, leading to a steady-state axisymmetric flow with no variabilities. As α increases, there is a threshold value at which $m\Omega_{\rm ISCO}$ type modes begin to appear. This threshold value of α is larger for larger c_{s0}. For example, the threshold is $\alpha \gtrsim 0.05$ for $c_{s0} = 0.01$ and $\alpha \gtrsim 0.1$ for $c_{s0} = 0.02$. Once this threshold is reached, the dominant $m\Omega_{\rm ISCO}$ mode initially involves large m, such as $m = 4$ seen in run (b). As α is further increased, the dominant m of these modes becomes smaller, for example $m = 2$ in run (c) and $m = 1$ in run (d). As α increases further, beyond a second threshold value, axisymmetric $\kappa_{\rm max}$ modes become present, while the $m\Omega_{\rm ISCO}$ modes may still be present with comparable power, as in run (d), or may be suppressed, such that only $\kappa_{\rm max}$ modes are present, as in runs (g) and (i). In between the two threshold values of α, it is possible for the $m\Omega_{\rm ISCO} \pm \kappa_{\rm max}$ modes to be present, sometimes in the absence of strong $\kappa_{\rm max}$ modes, such as in runs (c) and (f), probably due to a nonlinear coupling between the two other types of modes.

As reviewed in Section 2.2, Kato et al. (1988a) derived a criterion a for a viscous pulsational instability at the sonic point r_c, defined by $u_r(r_c) = c_s$, of a transonic flow. According to this criterion, the sonic point is unstable if $\alpha > u_c'/\Omega_c$, where u_c and Ω_c are the radial velocity and angular velocity at the critical point and the prime denotes a radial derivative. Note that these quantities themselves depend on α (and c_{s0}) in a self-consistent solution of the flow. While this

instability criterion may not be exactly applicable in our simulations (due to our 'diffusive' rather than 'αP' viscosity prescription), we find it is still a useful diagnostic. Using azimuthally-averaged and time-averaged profiles of u_r and u_ϕ, we measure the location of the sonic point r_c and the value of u_c'/Ω_c for each run, which are presented in Table 2.1. From these we can see that the Kato et al. criterion successfully indicates whether or not overstable oscillations are found, with the exception of run (*b*), for which a weak $m = 4$ mode is found even though the criterion is not strictly satisfied.

The results of our simulations of transonic disks can be most directly compared to those of Chan (2009), who also examined the role of viscosity in height-integrated disks (with a more general equation of state), and found evidence for viscous excitation of global axisymmetric and non-axisymmetric *p*-modes of the types found in our simulations. Our work extends Chan's analysis by delineating the different regimes (ranges of c_{s0} and α) in which particular combinations of these modes are excited, and highlights the possibility of extreme nonlinearities and the existence of the mixed $m\Omega_{ISCO} \pm \kappa_{max}$ type modes. The driving of the κ_{max} modes at large α is consistent with the 1D (height-integrated and axisymmetric) simulations of Milsom & Taam (1996) and Mao et al. (2009), as well as the 2D (axisymmetric) simulations of O'Neill et al. Note that the aforementioned studies are unable to capture the global spiral modes seen in our simulations due to their assumptions of axisymmetry.

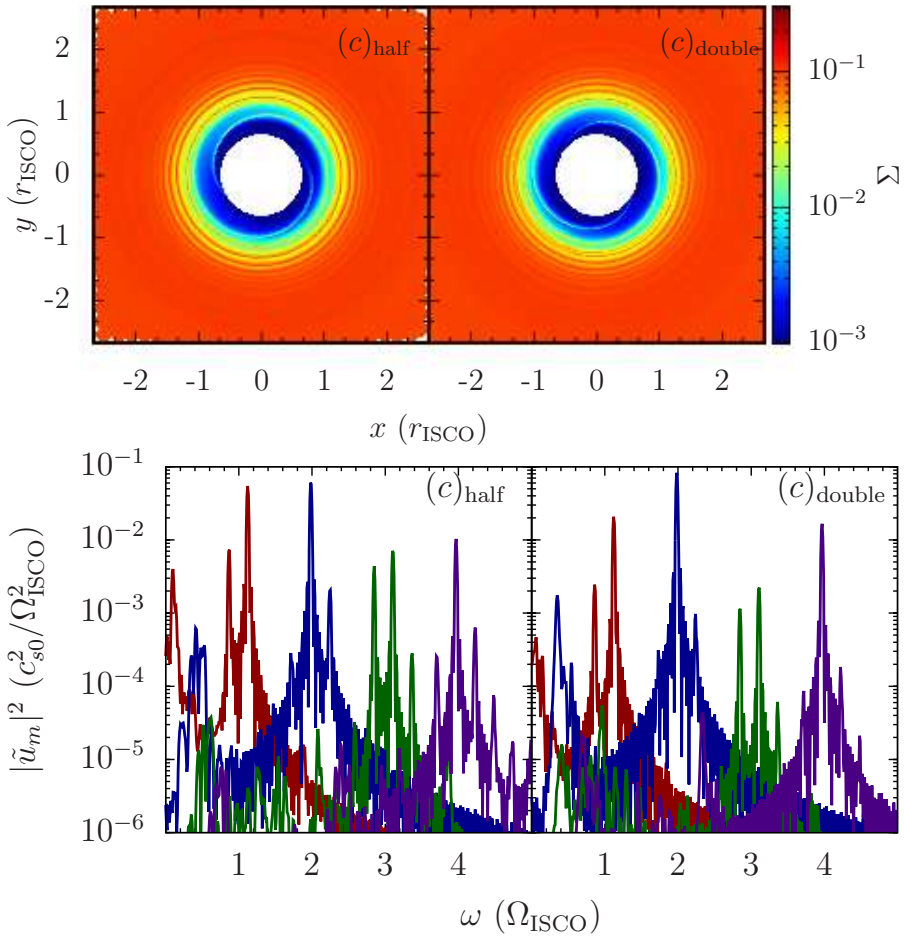

Figure 2.6: Surface density at $t = 100$ and power spectrum of resolution test runs $(c)_{\mathrm{half}}$ and $(c)_{\mathrm{double}}$ (compare to Figures 2.2 and 2.5).

2.4.3 Resolution Study

In order to determine the robustness and dependence on grid resolution of the features seen in our transonic disk simulations, we perform run (c) with both half the standard resolution in both dimensions, $N_r \times N_\phi = 512 \times 128$, denoted $(c)_{\text{half}}$ and double the standard resolution $N_r \times N_\phi = 2048 \times 512$, denoted $(c)_{\text{double}}$. The surface density at $t = 100$ and the m-delineated power spectra for these runs are shown in Figure 2.6. It can be seen from these that the runs are remarkably similar to one another, as well as to the standard resolution run, therefore we conclude that the observed features are not sensitive to resolution and that the standard resolution provides sufficient convergence. Since run (c) has the largest Reynolds number $\text{Re} = \alpha^{-1} c_{s0}^{-2}$ of the runs which exhibit all three types of oscillations, we conclude that our other runs are at least as well resolved as (c) under the standard resolution.

2.5 Overstable Global Oscillations in Truncated Disks

We now investigate the behavior of trapped p-modes in a disk truncated at its inner edge by an impermeable wall. As noted before, such an inner disk edge mimics a magnetospheric boundary. We place the inner boundary at r_{ISCO} and impose a reflecting boundary condition with $u_r = 0$ (imposed by anti-symmetrizing u_r across the boundary), and with all other variables fixed at their equilibrium values. In general, viscous forces induce a radial drift, which would lead to accumulation of mass at the inner boundary. We avoid this by initializing the surface density with a profile such that there is no radial drift (equation

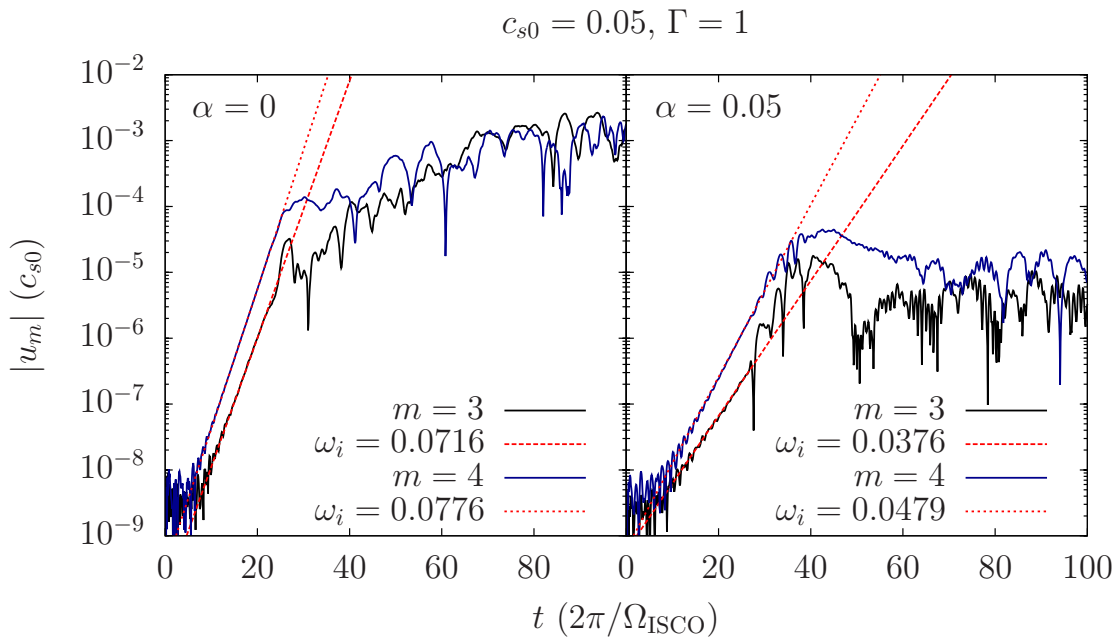

Figure 2.7: Evolution of the $m = 3$ and $m = 4$ components of radial velocity [$|u_m|$, in units of $c_{s0} = c_s\,(r_{\rm ISCO})$] at $r = 1.05$ for a typical run, showing exponential growth of trapped p-mode amplitudes, for zero and non-zero α. The dashed and dotted lines show the exponential curves used to fit the mode growth rates. The reduction in growth rate and reduced saturation amplitude due to viscosity can be seen.

	m	$(\omega_r, \omega_i)_{\text{semi-analytic}}$	ω_r	$\omega_{i,\alpha=0}$	$\omega_{i,\alpha=0.01}$	$\omega_{i,\alpha=0.05}$
$c_{s0} = 0.02, \Gamma = 1$	2	1.78, 0.0422	1.78	0.0453	0.0314	0.0266
	3	2.69, 0.0528	2.69	0.0579	0.0426	0.0235
	4	3.62, 0.0607	3.63	0.0646	0.0478	0.0205
$c_{s0} = 0.02, \Gamma = 4/3$	2	1.78, 0.0408	1.78	0.0434	–	–
	3	2.69, 0.0512	2.69	0.0550	–	–
	4	3.62, 0.0591	3.63	0.0611	–	–
$c_{s0} = 0.05, \Gamma = 1$	2	1.62, 0.0588	1.62	0.0600	0.0526	0.0211
	3	2.48, 0.0716	2.47	0.0716	0.0642	0.0376
	4	3.36, 0.0800	3.36	0.0776	0.0704	0.0479
$c_{s0} = 0.05, \Gamma = 4/3$	2	1.62, 0.0553	1.61	0.0552	–	–
	3	2.48, 0.0682	2.47	0.0665	–	–
	4	3.36, 0.0763	3.36	0.0724	–	–

Table 2.2: Summary of trapped p-mode simulations. From left to right, the columns give the equation of state (c_{s0} and Γ), azimuthal mode number m, the semi-analytic mode frequency ω_r and growth rate ω_i (zero viscosity), the numerically-calculated mode frequency, and the numerically-calculated growth rate for $\alpha = 0$ and several non-zero values of α. The blank entries correspond to runs for which p-modes were not observed due to a second type of instability, as described in Section 2.5.3.

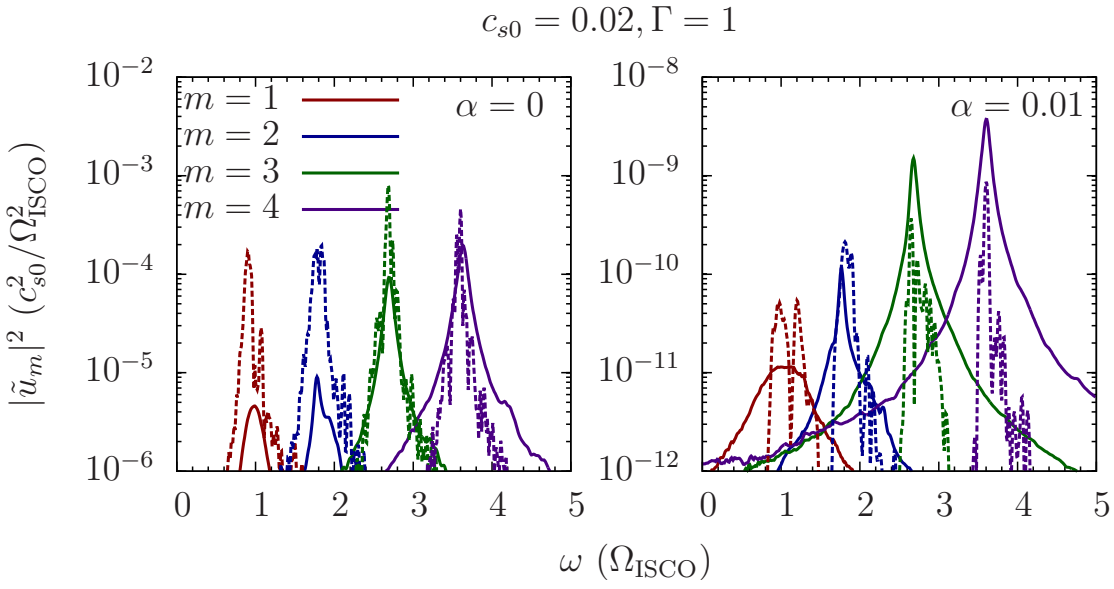

Figure 2.8: Power spectra delineated by azimuthal number m in the linear phase (solid curves) and nonlinear phase (dashed curves) for trapped p-modes with zero (left) and non-zero (right) α. The linear phase power spectra have been multiplied by 10^{10} and 10^4 (left and right, respectively) in order to compare them to the nonlinear phase power spectra.

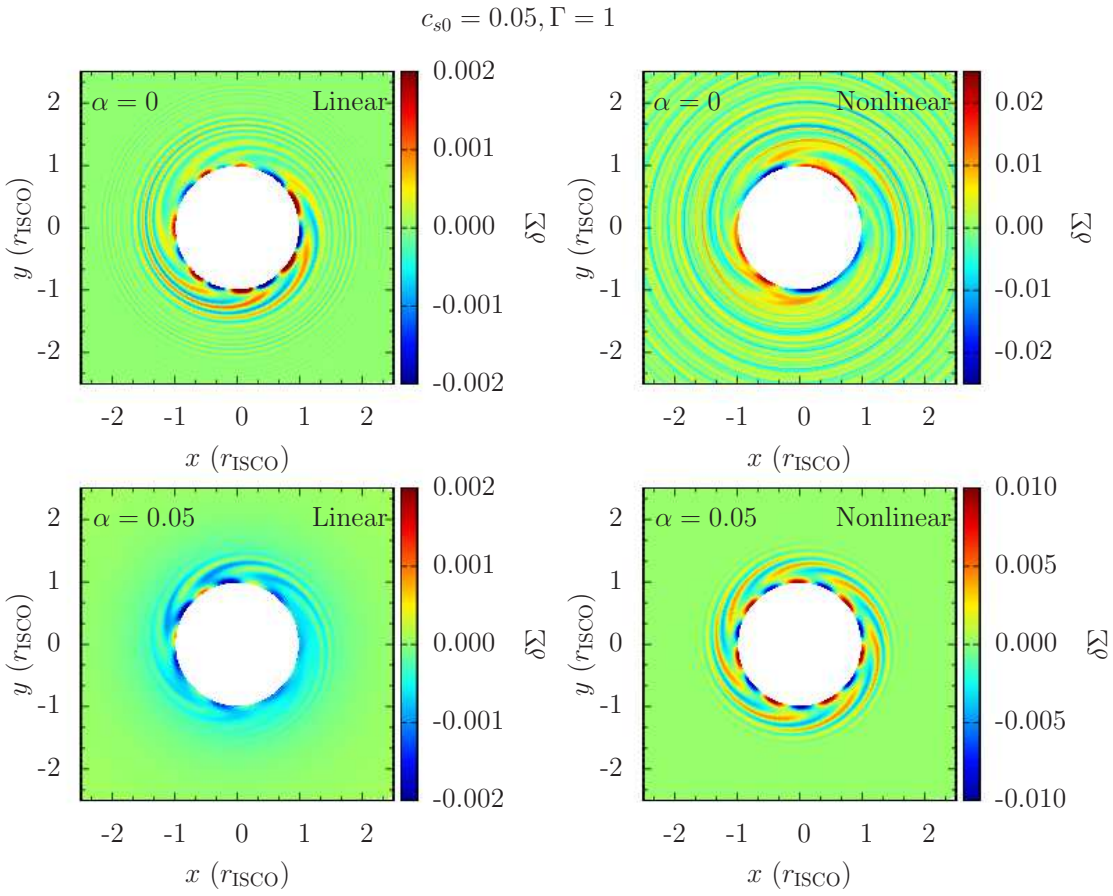

Figure 2.9: Surface density perturbations near the end of the linear phase (left) and in the nonlinear phase (right) with zero (top) and non-zero (bottom) α, showing the qualitative difference in saturation of trapped p-modes between the two cases.

2.8), given by

$$\Sigma(r) \propto \left[r^{-2} \left(\frac{r - r_s}{3r - r_s} \right) \right]^{1/\Gamma}. \tag{2.25}$$

This ensures that the azimuthally-averaged surface density does not change significantly as long as the perturbations to this equilibrium remain small (in principle this may be violated when the perturbation amplitudes become large and the modes are nonlinear). This initial surface density is modified by small ($\delta\Sigma/\Sigma \lesssim 10^{-6}$) random perturbations which serve as seeds for the growth of p-modes. They lead to the growth of many modes with many different values of m, the frequencies and growth rates of which can be measured independently from a single run (the accuracy of this approach is examined in Section 2.5.4).

Each mode is assumed to have a complex frequency $\omega = \omega_r + i\omega_i$ where ω_r and ω_i (both real) are the mode frequency and growth rate. The amplitude of radial velocity perturbations with azimuthal number m at radius r are determined numerically by computing

$$u_m(r,t) = \frac{1}{2\pi} \int_0^{2\pi} u_r(r,\phi,t)\, e^{-im\phi} \mathrm{d}\phi. \tag{2.26}$$

By fitting a straight line to $\log|u_m|$ in the linear growth phase, we can determine the growth rate of each mode. The frequencies are determined using a power spectrum delineated by azimuthal number m (equation 2.24). Some examples of analysis using these tools will be shown in the subsequent discussion of the results of the simulations.

Figure 2.7 shows the evolution of $|u_m|$ (for $m = 3$ and $m = 4$, evaluated at a representative radius) for a pair of typical runs. Initially there is a linear growth phase lasting 20 – 40 orbits, in which the amplitudes grow exponentially. The fitted curves used to measure the growth rates are shown. After the departure of the mode amplitudes from simple exponential growth, we refer to the rest of the

evolution as the nonlinear phase. Example power spectra are shown in Figure 2.8, showing the mode frequencies in both the linear and nonlinear phases.

2.5.1 Inviscid Case

We first perform simulations of an inviscid disk ($\alpha = 0$). Since the complex frequencies of the trapped p-modes can be compared to linear theory (Lai & Tsang 2009), these serve as test cases for verifying the robustness and accuracy of the simulations before we investigate the effects of viscosity (see Fu & Lai 2013 for a more detailed analysis of numerical simulations of trapped p-modes in inviscid disks). We choose four equations of state, with $c_{s0} = 0.02$ or $c_{s0} = 0.05$ and $\Gamma = 1$ or $\Gamma = 4/3$. The main results, numerically measured frequencies and growth rates of the modes, are presented in Table 2.2. For runs with $c_{s0} = 0.02$, the numerical growth rates agree with those computed from linear theory to better than 10%, while for $c_{s0} = 0.05$ they agree to better than 5%, with some cases in even better agreement. In all cases the numerical mode frequencies agree with linear theory to better than 1%, and these frequencies differ very little between the linear and nonlinear phases (see Figure 2.8). See Section 2.5.4 for a resolution test demonstrating that these results are converged. Having accurately captured the properties of the inviscid trapped p-modes, we move on to exploring the effects of viscosity in the next subsection.

2.5.2 Viscous Effects

We carry out simulations using the same parameters described above but now with non-zero viscosity, choosing $\alpha = 0.01$ and $\alpha = 0.05$. Our analysis will focus on the modification of mode growth and saturation properties due to viscous effects. We separate this discussion into two parts based on the value of the polytropic index, as the effects of viscosity are different for $\Gamma = 1$ and $\Gamma = 4/3$.

The effect of viscosity on the mode growth rates are summarized in Table 2.2. We find qualitatively different results depending on the value of the polytropic index Γ. For $\Gamma = 1$, the growth rates of all modes are reduced compared to the inviscid case. For $c_{s0} = 0.05$, the difference in growth rate due to viscosity $\delta\omega_\nu = \omega_i - \omega_i^{(\alpha=0)}$, is approximately proportional to α, as expected theoretically (equation 2.12). This is not the case for $c_{s0} = 0.02$, for which such an extrapolation of ω_i based on the $\alpha = 0.01$ case would imply a negative value of ω_i for $\alpha = 0.05$ (and thus the modes would not be overstable), which is not what we find. More importantly, the linear analysis of viscous effects predicts enhanced growth rates, rather than the diminished ones measured numerically. This indicates that the estimate of the viscous effect on the growth of global modes (equation 2.14) is inaccurate, perhaps due to dynamics of the corotation region which are neglected in equation (2.14).

The saturation behavior of the the modes in the nonlinear phase is also affected by viscosity. Figure 2.7 shows that while in the inviscid case, the various u_m continue to grow slowly after the linear phase before saturating at larger amplitudes, with viscosity they decrease or level off after the linear phase, remaining somewhat steady at a relatively small amplitude. This effect can also be seen in Figure 2.9, which shows that when viscosity is included, the disk

perturbations remain remarkably similar to the linear phase in appearance during the nonlinear phase, unlike the inviscid case for which the nonlinear phase looks qualitatively different than the linear phase. Therefore it appears that viscous forces ultimately suppress the mode amplitudes enough to prevent any significant nonlinear effects from occurring.

Our simulations with reflective inner boundary yield unexpected results for $\Gamma = 4/3$. Unlike the case of $\Gamma = 1$ for which the overstable p-modes behave similarly with non-zero α, albeit with reduced growth rates, when $\Gamma = 4/3$ we find no trace of the trapped p-modes. Instead, the power spectrum for these runs shows oscillations with frequencies approximately equal to $m\Omega_{\rm ISCO}$. For these runs, the evolution of $|u_m|$ does not show clean exponential growth as the p-modes do, but rather a sharp initial rise to a significant amplitude followed by slow incoherent growth. The amplitude of these oscillations eventually becomes unphysically large with radial velocities much greater than $c_{\rm s0}$. As the behavior of these unexpected modes is drastically different from the original p-modes, we forego detailed analysis of these runs. This highly unstable unexpected behavior may be the result of the combined effects of overstability due to GR (resulting from the smallness of κ/Ω and increased q) and due to viscosity (resulting from large A), both of which enhance the growth rate as given in eq. (2.12). Since we are interested in the viscous effect, in Section 2.5.3 we will isolate it by modifying the numerical setup to simulate a disk with a pure Newtonian $1/r$ potential, to see if this behavior persists in a simple way that allows for a concrete analysis.

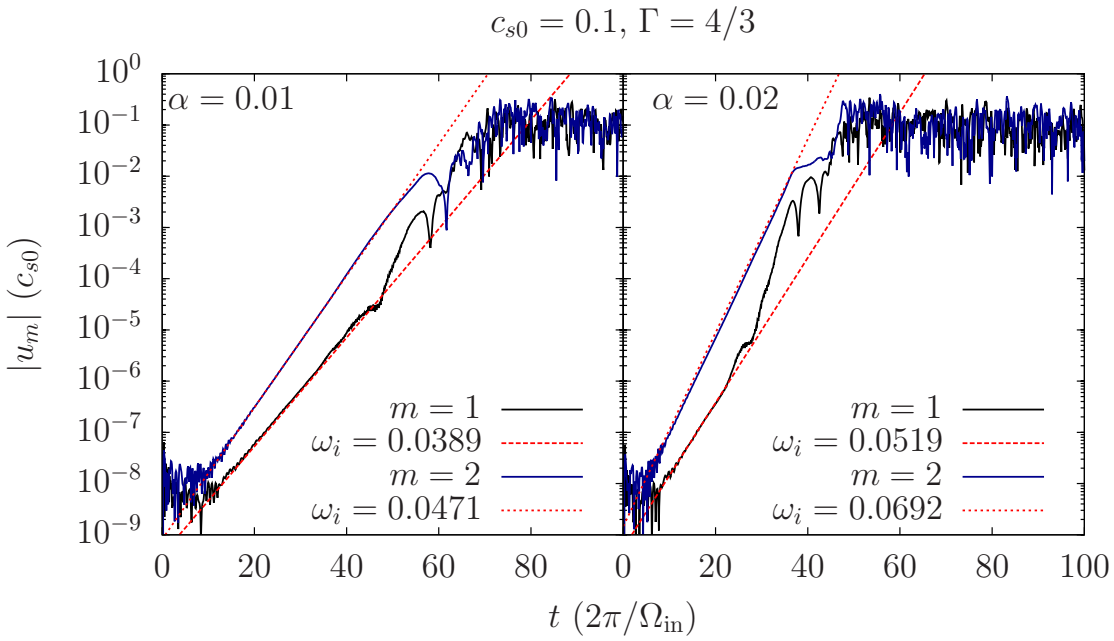

Figure 2.10: Evolution of $|u_m|$ at $r = 1.05$ for Newtonian disk with reflective inner boundary. The $m = 1$ and $m = 2$ components are shown, along with exponential fits used to determine growth rates, for two values of α. The growth rates increase with increasing α, in qualitative agreement with the theory of viscous overstability.

2.5.3 Viscosity-Driven Boundary Modes in Newtonian Potential

We wish to determine whether or not the growth of the new modes which overwhelm the p-modes is due to a purely viscous effect which can operate in the absence of GR effects. To this end, we perform simulations using the same setup as described at the beginning of this section, but replace the pseudo-Newtonian gravitational potential with a pure Newtonian potential, $\Phi = -GM/r$. Under this potential there is no longer an ISCO, so the location of the inner boundary no longer has any physical significance, and we simply denote it as $r_{\rm in}$. The corresponding boundary condition, as in the beginning of this section, is

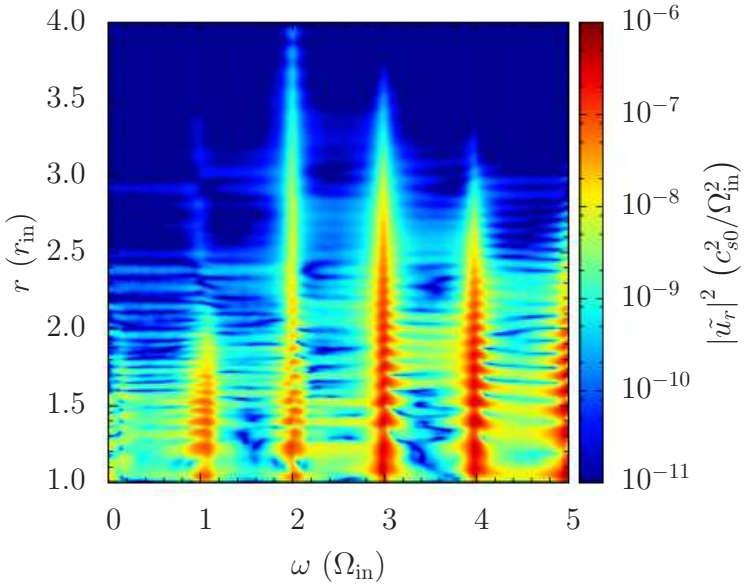

Figure 2.11: Linear phase power spectrum for viscosity-driven boundary modes in Newtonian disk.

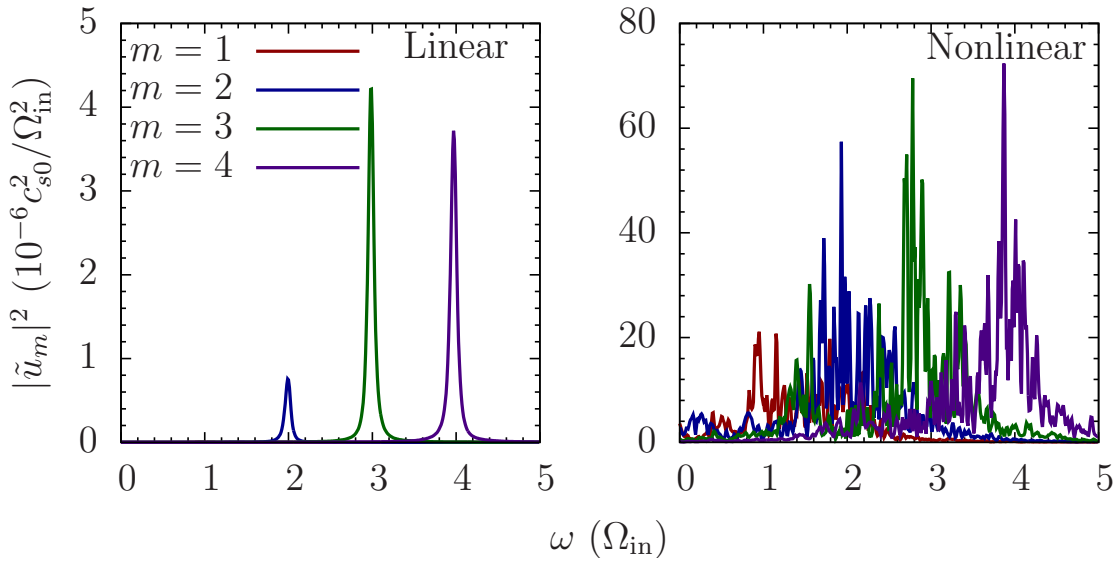

Figure 2.12: Power spectra for lowest m boundary modes in linear and nonlinear phases. In the linear phase, each mode corresponds to a single sharp frequency peak close to $m\Omega_{\rm ISCO}$, while in the nonlinear phase these peaks are broader and exhibit splittings, but retain powers concentrated in frequencies similar to those found in the linear phase.

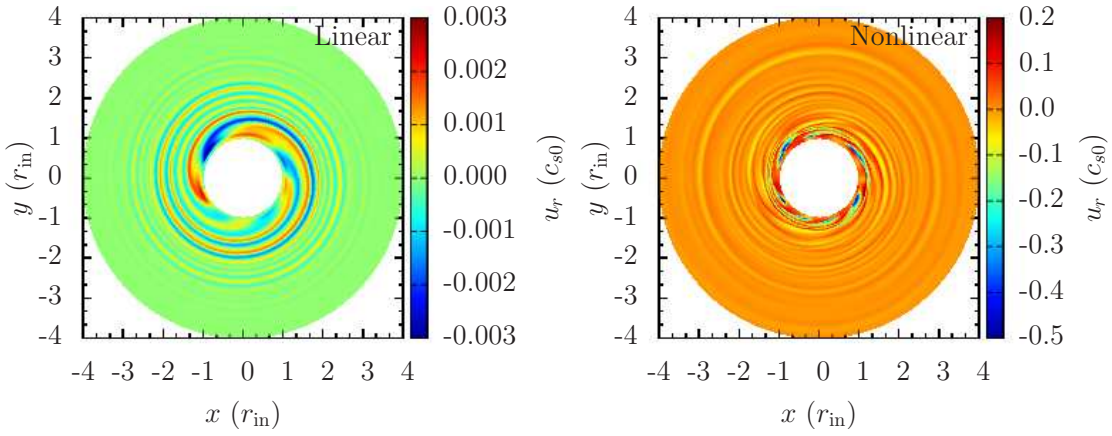

Figure 2.13: Radial velocity of boundary modes in Newtonian disks in the linear and nonlinear phase.

	m	$\omega_{\rm r}$	$\omega_{\rm i}$
$\alpha = 0.01$	1	1.05	0.0389
	2	2.00	0.0471
	3	2.99	0.0488
	4	3.98	0.0488
$\alpha = 0.02$	1	1.01	0.0519
	2	2.02	0.0692
	3	3.01	0.0743
	3	4.00	0.0744

Table 2.3: Frequencies and growth rates of boundary modes.

reflective with $u_r = 0$. We choose four sets of parameters, using $c_{s0} = 0.1$ and $(\Gamma, \alpha) = (1, 0), (1, 0.01), (4/3, 0), (4/3, 0.01)$. Due to the larger sound speed used, and therefore larger radial wavelengths, we reduce our resolution to $N_r \times N_\phi = 512 \times 128$.

We observe no growing oscillations for any of these runs except for $(\Gamma, \alpha) = (4/3, 0.01)$, confirming that the anomalous modes observing in Section 2.5.2 can be attributed purely to viscous effects (that is, requires $\alpha > 0$ and $\Gamma > 1$) and do not depend on GR effects. We therefore performed an additional run with $(\Gamma, \alpha) = (4/3, 0.02)$ to probe their dependence on α. Figure 2.10 shows the evolution of the lowest m mode amplitudes (represented by $|u_m|$ at $r = 1.05$), showing their exponential growth at a rate which increases with α (see Table 2.3 for complete list of frequencies and growth rates), as well as their relatively flat saturation amplitude consistent with the the viscous effect on mode saturation seen in Section 2.5.2. Figure 2.11 shows the power spectrum for these modes, indicating that they are global, with power in discrete frequencies across a range of radii. This justifies the use of eq. (2.24) to produce a radially integrated m-delineated power spectrum, as shown in Figure 2.12. From this we see that in the nonlinear phase, the modes retain approximately the same frequencies as in the linear phase, but with broader power spectra peaks and the presence of other sub-dominant peaks at new frequencies. The morphology of the two phases is shown in Figure 2.13. Note that the linear phase bears some resemblance to that of the p-modes (compare to Figure 2.9), while the nonlinear phase appears very different, having a more complex appearance.

These modes may be related to the interface modes studied by Tsang & Lai (2009b) and Fu & Lai (2012), who considered a Keplerian disk in contact with a

α	$(\omega_r, \omega_i)_{\text{half}}$	$(\omega_r, \omega_i)_{\text{standard}}$	$(\omega_r, \omega_i)_{\text{double}}$
0	$1.79, 0.0396$	$1.78, 0.0419$	$1.78, 0.0423$
0.01	$1.77, 0.0301$	$1.76, 0.0315$	$1.76, 0.0317$

Table 2.4: Frequency and growth rate of $m = 2$ trapped p-mode for $(c_{s0}, \Gamma) = (0.02, 1)$ with initial $m = 2$ perturbation for half, standard and double resolution runs.

uniformly rotating cylinder (for example, a magnetosphere), which is similar to the setup we consider. Their results indicate that modes with frequencies close to $m\Omega_{\text{in}}$ can be overstable due to the corotation resonance effect. In our case, they may not be intrinsically overstable but are only driven by viscous forces which can act in phase with the oscillations, by having a sufficiently large A parameter. A qualitative description of this phenomenon is as follows. In our simulation, the disk is initialized with a surface density for which viscous forces induce zero radial drift. However, a small perturbation to this equilibrium leads to a non-zero u_r, which leads to a radial oscillation at the local epicyclic frequency (which is simply equal to the orbital frequency Ω in a Newtonian potential). When material close to the inner solid boundary is pushed inward due to this oscillation, it is compressed, leading to a stronger viscous forces, which act as an additional restoring force (in addition to gravity which is producing the epicyclic motion). In this way, the restoring force increases with the amplitude of the oscillation, leading to overstability.

2.5.4 Resolution Study

As in Section 2.4.3, we perform two resolution test runs with half (512×128) and double (2048×512) the standard resolution, for the case $(c_{s0}, \Gamma, \alpha) = (0.02, 1, 0)$. These runs also differ from the standard run in that the initial density pertur-

bations are proportional to $\cos(2\phi)$, so that the $m = 2$ mode is isolated. This allows us to test the accuracy of measuring mode frequencies and growth rates for modes of multiple m from an initially random perturbation, as is done in our standard runs. The results are shown in Table 2.4. We see that with increasing resolution, the numerically measured growth rate converges on the result from linear theory, agreeing to better than 1% for the double resolution run. Additionally, we see that a standard resolution run with a pure $m = 2$ initial perturbation reproduces the growth rate of this mode more accurately than a run with random perturbations, as used in Section 2.5.2. However, the difference is small enough to justify the use of the random perturbation approach, which allows us to capture multiple modes from single runs. We note that since this resolution test was performed for our smallest sound speed, for which the modes have the smallest radial wavelengths, these tests provide a lower limit on our accuracy, as we expect the longer wavelength modes associated with our larger sound speed to be better resolved by the same resolution. We also perform the same test using $(c_{s0}, \Gamma, \alpha) = (0.02, 1, 0)$ to ensure that viscous effects are not sensitive to resolution. For this case we achieve similarly excellent convergence of growth rate with increasing resolution. Additionally, the discrepancy between the run with random perturbations and the one with $m = 2$ perturbations is smaller than the inviscid case.

2.6 Discussion

We have carried out viscous hydrodynamical simulations of black hole (BH) accretion disks to determine how viscosity may drive global oscillation modes in the disk. Analytical considerations (see Section 2.2) indicate that, depend-

ing on the viscosity law (particularly how viscosity scales with density) and the background disk flow properties, various global oscillations with different azimuthal wavenumbers may be excited. We considered two types of accretion flows. The first involves disks with transonic radial inflow across the innermost stable circular orbit (ISCO) of the BH, and the second involves disks truncated at an inner edge and with a free outer boundary at large distance.

Our simulations on transonic disks (Section 2.4) follow previous works in one dimension (Milsom & Taam 1996) and in two or three dimensions (O'Neill et al. 2009; Chan 2009). We show that, depending on the viscosity parameter α, global oscillations with different azimuthal wavenumbers m, and with frequencies close to $\kappa_{\rm max}$, $m\Omega_{\rm ISCO}$ or their linear combinations (where $\kappa_{\rm max}$ and $\Omega_{\rm ISCO}$ are the maximum radial epicyclic frequency and the orbital frequency at the ISCO, respectively) are excited. In general, as α increases above a critical value $\alpha_{\rm crit1}$ ($\sim 0.1 - 0.25$ for isothermal disks), non-axisymmetric modes (typically $m = 4$) first develop, then axisymmetric modes develop above a second critical value $\alpha_{\rm crit2}$ (~ 0.5). We propose that these modes are the result of viscous overstability, possibly combined with an effect related to instability of the sonic point, although the theory of the latter is not strictly applicable in our simulations.

Our simulations on truncated disks (Section 2.5) extend previous semi-analytic works on overstable inertial-acoustic modes (p-modes) driven by corotation resonance in non-magnetic disks (Lai & Tsang 2009; Tsang & Lai 2008, 2009b; Horak & Lai 2013) and magnetic disks (Fu & Lai 2009; 2011), on magnetosphere-disk interface modes (Tsang & Lai 2009a; Fu & Lai 2012), and numerical studies of nonlinear p-modes (Fu & Lai 2013). We find that the growth rates of overstable p-modes are reduced by viscosity in isothermal disks,

while for other equations of state (and corresponding viscosity laws), they are suppressed due to the excitation of a different class of modes with frequencies $\omega \approx m\Omega_{\rm in}$, which are related to magnetosphere-disk interface modes.

Obviously, our 2D hydrodynamical disk models do not capture various complexities (e.g., magnetic field, turbulence and radiation) associated with real BH accretion disks. Much progress in numerical General Relativistic Magnetohydrodynamic (GRMHD) simulations of BH accretion disks has been made in the past decade, but much work remains to understand the complex phenomenology of BH X-ray binaries (e.g., Remillard & McClintock 2006; Done et al. 2007; Belloni et al. 2012). Several recent simulations have revealed quasi-periodic variabilities of various fluid variables, but the connection of these variabilities to the observed HFQPOs is far from clear (e.g., Henisey et al. 2009; ONeill et al. 2011; Dolence et al. 2012; McKinney et al. 2012; Shcherbakov & McKinney 2013). Our simulations, although based on highly simplified disk models, demonstrate that under appropriate conditions, various global oscillation modes can grow to nonlinear amplitudes with well-defined frequencies. We emphasize that our height-integrated treatment cannot capture modes with dependence on vertical structure (g-modes and c-modes), which may be present in 3D simulations. The properties and excitation of the p-modes considered in this paper may also be modified in a fully 3D approach.

It is of interest to compare the frequencies of our simulated modes to those of observed HFQPOs in X-ray binaries (Remillard & McClintock 2006; Belloni et al. 2012), particulary the three systems whose BH spins are constrained using the continuum-fitting method (see Narayan & McClintock 2012 and references therein). These considerations indicate that the frequencies of the HFQPOs are

smaller than the ISCO frequency, therefore they are unlikely to be explained by any non-axisymmetric modes with $\omega \gtrsim \Omega_{\rm ISCO}$. However, this does not rule out the axisymmetric $\kappa_{\rm max}$ modes which have lower frequencies. We note that we have not addressed the mechanism by which disk oscillations may manifest as variabilities in X-ray flux, which affects the observability (or lack of observability) of any oscillations that occur in the disk.

Although our simulations cannot be directly compared with observations, some tentative conclusions can be drawn. The thermal (high-soft) state of BH X-ray binaries may be approximately described by the transonic disk model (e.g. Done et al. 2007). Our simulations show that such a disk does not allow trapping of p-modes, but can excite global $\kappa_{\rm max}$-waves or $m\Omega_{\rm ISCO}$-waves when the viscosity parameter α is sufficiently large. Since no significant HFQPOs are observed in the thermal state of X-ray binaries, our simulations provide an upper limit on the effective disk viscosity parameter of $\alpha \lesssim 0.5$.

The physical nature of the intermediate state (or "steep power law state") of BH X-ray binaries, when HFQPOs are observed, is currently uncertain (Done et al. 2007; Oda et al. 2010). Since episodic jets are produced in this state, large-scale magnetic fields likely play an important role (e.g., Tagger & Varniere 2006; Yuan et al. 2009; McKinney et al. 2012). When magnetic fields advect inwards in the accretion disk and accumulate around the BH (e.g. Bisnovatyi-Kogan & Ruzmaikin 1974, 1976; Igumenshchev et al. 2003; Rothstein & Lovelace 2008), a magnetosphere may form. The truncated disk model studied in this paper may mimic the disk outside the magnetosphere. Our calculations show that overstable p-modes (driven by corotation resonance) can be trapped in the inner disk, but their growth rates are typically reduced by disk viscosity. More

importantly, with increasing α and more general equation of state, a new series of global waves can be excited at the disk boundary (see Tsang & Lai 2009a; Fu & Lai 2012; McKinney et al. 2012).

CHAPTER 3

ROSSBY WAVE INSTABILITY AND LONG-TERM EVOLUTION OF DEAD ZONES IN PROTOPLANETARY DISKS

3.1 Introduction

The transport of angular momentum is a central problem in the evolution of protoplanetary disks. The magnetorotational instability (MRI; Balbus & Hawley 1991, 1998) in magnetized Keplerian disks can sustain turbulence which effectively transports angular momentum, provided that the disk material is sufficiently ionized. This condition is thought to be met in the inner disk ($\lesssim 0.1 - 1.0$ AU), due to thermal ionization, and in the outer disk (a few to $\gtrsim$ 10's of AU), due to irradiation by high energy photons and cosmic rays. At intermediate radii, the disk is cold and and weakly ionized, producing an MRI-inactive "dead zone" (DZ) in which turbulence is significantly suppressed in the disk midplane, although thin surface layers may remain turbulent (Gammie 1996; see Armitage 2011 for a review). The spatial extent of the DZ and its level of residual turbulence depend on many factors. These include the ionization fraction, which is affected by the abundance of dust grains (e.g., Sano et al. 2000; Desch & Turner 2015) and shielding of cosmic rays by stellar winds (Cleeves et al. 2013), the role of non-ideal MHD effects, such as Ohmic diffusion, ambipolar diffusion, and Hall drift (Bai & Stone 2013; Bai 2013, 2014a, 2014b; Lesur et al. 2014; Simon et al. 2015; see Turner et al. 2014 for a review), and the direct interaction of the magnetized stellar wind with disk surface layers (Russo & Thompson

This chapter is adapted from Miranda, Lai, & Méheut (2016).

2015).

In the absence of MRI-driven turbulence, one or more alternative angular momentum transport processes must operate in order to allow protoplanetary disks to evolve in accordance with their typical observed lifetimes and accretion rates (a few Myr and $10^{-8}\,M_{\odot}$/yr, e.g., Hartmann et al. 1998; Haisch, Lada & Lada 2001). Several such mechanisms have been proposed, such as gravitational instability (e.g., Lodato & Rice 2004; Rafikov 2015), disk winds (e.g., Bai & Stone 2013), baroclinic instability (e.g., Klahr & Bodenheimer 2003), and vertical shear instability (e.g., Urpin & Brandenburg 1998). These hydrodynamical instabilities may produce locally correlated velocity fluctuations, leading to enhanced effective viscosity (Balbus & Papaloizou 1999). In these cases, the transport of angular momentum does not necessarily result from turbulence, but instead may be due to large-scale structures.

A promising mechanism for reviving the DZ is the Rossby wave instability (RWI), a global, non-axisymmetric instability which arises at "bumps" in Keplerian disks (Lovelace et al. 1999; Li et al. 2000; Méheut et al. 2010, 2012a, 2012b, 2013; see review by Lovelace & Romanova 2014). More specifically, RWI is associated with narrow radial minima of vortensity (vorticity divided by surface density), and leads to growth of vortices, which then merge into a single vortex (Li et al. 2001). The anticyclonic rotation (i.e., opposite to the bulk disk rotation) of the resulting vortex prevents its destruction by Keplerian shear, allowing it to potentially have a long lifetime (e.g., Godon & Livio 1999).

The sharp gradients in viscosity at DZ edges[1] naturally produce RWI-

[1]When angular momentum transport in the the DZ is suppressed by Ohmic diffusion (e.g., Gammie 1996; Fleming & Stone 2003), the abrupt onset of MRI at a threshold resistivity level ensures that the gradient of effective viscosity is sharp, even when that of the underlying ionization/resistivity is not (Lyra et al. 2015). It is unclear whether or not the DZ has a similarly

unstable vortensity profiles. Vortices and associated spiral density waves produced by the RWI create fluid stresses that transport angular momentum, which can revive the DZ (Varnière & Tagger 2006; Lyra et al. 2009b; Lyra & Mac Low 2012; Regály et al. 2012). Further, the low viscosity DZ is a favorable location for the production of vortices, since viscosity can inhibit the RWI (Lin 2014; Gholipour & Nejad-Asghar 2014), and influence the long-term survival of vortices (Fu et al. 2014a; Zhu & Stone 2014). In this paper, we examine the role of RWI and vortices in the evolution of disks with DZs, building on existing work by considering both the inner and outer DZ edges together (especially in the case when they are close to each other), and by focusing on the long-term, quasi-steady behavior of the disk.

Anticylonic vortices, such as those produced by RWI, may play a role in planet formation due to their ability to trap dust particles (Barge & Sommeria 1995; Godon & Livio 2000; Tanga et al. 1996; Lyra et al. 2009b; Méheut et al. 2012c; Lyra & Lin 2013; Zhu & Stone 2014), which may produce the conditions needed for the formation of planetesimals (see review by Chiang & Youdin 2010). However, feedback on the gas by the accumulated dust particles may subsequently destroy the vortex (e.g., Chang & Oishi 2010; Méheut et al. 2012c; Fu et al. 2014b). Recent observations of transitions disks (disks with central holes in dust emission, see review by Espaillat et al. 2014) show strong asymmetries in mm-dust emission (e.g., van der Marel et al. 2013; Casassus et al. 2013; Isella et al. 2013), which could possibly be explained by dust trapping in vortices. These vortices may be produced at DZ edges (e.g., Regály et al. 2012), or at the edges of gaps opened by planets (e.g., Zhu & Stone 2014). In the simulations presented in this paper, novel asymmetries are produced by the

sharp edge when other non-ideal MHD effects or disk winds play a significant role.

gas dynamics associated with the presence of a DZ. Since we do not follow the dynamics of dust particles, which are coupled to the gas through aerodynamic drag, we cannot make concrete predictions about the resultant dust morphologies. However, some qualitative features can be extrapolated to the behavior of dust, such as azimuthal symmetries, and the location of features relative to the position of the DZ.

It has been suggested that the presence of a DZ may lead to episodic accretion. In this scenario, accumulation of mass in the DZ triggers local gravitational instability driven turbulence, which heats the DZ and briefly triggers MRI, causing an accretion outburst, before cooling and repeating the cycle (e.g., Armitage et al. 2001; Zhu et al. 2010a, 2010b; Martin & Lubow 2011, 2014). This mechanism has been used to explain the outburst behavior of FU Orionis systems. The majority of work on this topic has used one-dimensional models (with some exceptions, e.g., Bae et al. 2014), which neglect non-axisymmetric effects. In this work, we show that, as long as the accretion rate is not too high, vortices produced by the RWI (which is explicitly non-axisymmetric) can generate fluid stresses which facilitate steady, non-episodic accretion through the DZ. We place a limit on the factor by which the gravitational stability parameter (Toomre Q) is reduced in the DZ compared to a steady-state disk with no DZ.

In this paper, we give considerable attention to narrow DZs, for which the radial extent of the DZ is smaller than its distance from the central star. As we show, this configuration leads to large fluid stresses and produces novel morphologies, including unusual vortex shapes and azimuthal symmetries (mode numbers other than $m = 1$). These features are the result of coherent oscillations of the entire DZ. By contrast, for wide DZs, each edge behaves independently,

and the RWI evolves toward the familiar $m = 1$ symmetry (e.g., Méheut et al. 2012b), producing a much smaller Reynolds stress in the DZ. Estimates of the extent of the DZ in realistic disks are highly uncertain, and range from very narrow, as described here, to highly extended. Viscosity profiles resembling our parameterized DZs (i.e., regions with reduced viscosity relative to their surroundings) may be present near or in between ice lines, where there are significant changes in dust grain abundance, on which the strength of MRI turbulence depends sensitively (e.g., Kretke & Lin 2007; Bitsch et al. 2014). Therefore, there are plausible physical conditions in protoplanetary disks which may produce the DZ configurations modeled in this paper.

The main goal of this paper is to study the evolution of DZs using global, long-term hydrodynamic simulations. We find three main results. Disks with DZs evolve toward quasi-steady states, in which steady, non-episodic accretion is partially facilitated by fluid stresses produced by vortices and spiral density waves resulting from the RWI. We show that narrow DZs produce large Reynolds stresses, with only moderate surface density enhancements in the DZ, so they are unlikely to experience gravitational instability unless the accretion rate is high. We also present novel gas disk morphologies produced by the RWI in DZs.

The outline of this paper is as follows. In Section 3.2, we describe the the setup of our simulations. In Section 3.3, we present a detailed analysis of a canonical run, which demonstrates our main results. We present a suite of simulations which systematically explore the effects of varying the DZ parameters in Section 3.4. Finally, we discuss these results and conclude in Section 3.5. In the Appendix, we present a linear calculation of how viscosity affects the RWI.

3.2 Setup

We consider a 2D (height-integrated) disk described by surface density Σ and velocity $\boldsymbol{u} = \left(u_r, u_\phi\right)$, with a radial extent of $r_{\rm in} = 1$ to $r_{\rm out} = 12$. It experiences a gravitational potential (per unit mass) $\Phi = -GM_*/r$, where M_* is the mass of the central star, so the Keplerian orbital frequency is $\Omega_{\rm K} = (GM_*/r^3)^{1/2}$. We adopt a locally isothermal equation of state $P = c_{\rm s}^2(r)\,\Sigma$, where P is the height-integrated pressure, and

$$c_{\rm s}(r) = c_{\rm s0}\left(\frac{r}{r_{\rm in}}\right)^{-1/2} \tag{3.1}$$

is the radially-dependent sound speed. The scale height of the disk, $H = c_{\rm s}/\Omega_{\rm K}$, is proportional to r, so that the aspect ratio $h = H/r$ is constant. The kinematic viscosity is prescribed by

$$\nu = \alpha c_{\rm s} H. \tag{3.2}$$

The dimensionless viscosity parameter α is a function of r, given by

$$\alpha = \alpha_0 f_\alpha(r), \tag{3.3}$$

where $f_\alpha(r)$, which describes the shape of the DZ, is given by

$$f_\alpha(r) = 1 + \frac{1}{2}(1-\epsilon_{\rm DZ})\left[\tanh\left(\frac{r-r_{\rm ODZ}}{\Delta r_{\rm ODZ}}\right) - \tanh\left(\frac{r-r_{\rm IDZ}}{\Delta r_{\rm IDZ}}\right)\right]. \tag{3.4}$$

The meaning of the parameters are as follows: $r_{\rm IDZ}$ and $r_{\rm ODZ}$ are the inner and outer edges of the DZ, α_0 is the value of α in the active zones ($r < r_{\rm IDZ}$ and $r > r_{\rm ODZ}$), $\epsilon_{\rm DZ}$ is the factor by which α is reduced in the DZ (i.e., $\alpha = \epsilon_{\rm DZ}\alpha_0$ in the DZ), $\Delta r_{\rm IDZ}$ and $\Delta r_{\rm ODZ}$ are the widths of viscosity transitions at the inner and outer DZ edges. These, along with the aspect ratio h, fully describe each simulation. Unless otherwise stated, time is expressed in units of the Keplerian orbital period at $r_{\rm in}$, $P_{\rm in} = 2\pi/\Omega_{\rm in}$.

3.2.1 Initial Conditions

The initial surface density (in code units) is

$$\Sigma = r^{-1/2}, \tag{3.5}$$

which satisfies the steady state condition $\Sigma\nu$ = constant in the active zones ($\alpha = \alpha_0$), but not in the DZ, which would require Σ to be larger by a factor of $\epsilon_{\rm DZ}^{-1}$. The disk is initially in centrifugal balance with $u_\phi = r\Omega_0$, where

$$\Omega_0^2 = \Omega_{\rm K}^2 + \frac{c_{\rm s}^2}{r^2}\left(\frac{{\rm d}\ln\Sigma}{{\rm d}\ln r} + \frac{{\rm d}\ln c_{\rm s}^2}{{\rm d}\ln r}\right), \tag{3.6}$$

and initially $u_r = 0$ everywhere. Non-axisymmetric instabilities are seeded by random perturbations (i.e., they have no preferred azimuthal symmetry) to Σ, with amplitude $\delta\Sigma/\Sigma \lesssim 10^{-6}$.

3.2.2 Boundary Conditions

At the outer boundary ($r_{\rm out} = 12$), the fluid variables are fixed at the constant values

$$\Sigma = r^{-1/2}, \quad u_r = -\frac{3}{2}\frac{\nu}{r}, \quad u_\phi = r\Omega_0, \tag{3.7}$$

so that the accretion rate,

$$\dot{M} = -\int_0^{2\pi} \Sigma u_r r {\rm d}\phi, \tag{3.8}$$

supplied to the disk has the prescribed value $\dot{M}_0$. At the inner boundary ($r_{\rm in} = 1$), zero-gradient conditions are imposed on Σ and u_r, while u_ϕ is fixed at its initial (modified) Keplerian value, i.e.,

$$\frac{\partial\Sigma}{\partial r} = \frac{\partial u_r}{\partial r} = 0, \quad u_\phi = r\Omega_0. \tag{3.9}$$

Near the inner and outer boundaries, wave damping zones are used to minimize wave reflection. This is achieved by solving the following equations for the variables $\mathbf{x} = \left(\Sigma, u_r, u_\phi\right)$ at the end of each time step. In the outer damping zone $10 < r < 12$, they are damped to an equilibrium state,

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\frac{\mathbf{x} - \mathbf{x}_0}{\tau_{\mathrm{out}}} R\left(r\right), \tag{3.10}$$

where $\mathbf{x}_0 = \left(r^{-1/2}, -3\nu/2r, r\Omega_0\right)$, so that the accretion rate is relaxed to $\dot{M}_0$. In the inner damping zone ($1 < r < 2$), they are damped to their azimuthal averages,

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\frac{\mathbf{x} - \langle\mathbf{x}\rangle_\phi}{\tau_{\mathrm{in}}} R\left(r\right). \tag{3.11}$$

We choose the damping timescales $\tau_{\mathrm{in}} = 2\pi\Omega_{\mathrm{in}}^{-1}$ and $\tau_{\mathrm{out}} = 2\pi\Omega_{\mathrm{out}}^{-1}/10$, and the dimensionless envelope function is

$$R\left(r\right) = 1 - \left(\frac{r - r_{\mathrm{b}}}{r_{\mathrm{d}} - r_{\mathrm{b}}}\right)^2, \tag{3.12}$$

where r_{b} is the (inner or outer) boundary and r_{d} is where the corresponding damping zone begins. We verified that our numerical results are not strongly affected by the placement of the boundaries and damping zones.

3.2.3 Numerical Method

The fluid equations are solved using the finite volume, Godunov scheme hydrodynamics code PLUTO (Mignone et al. 2007). We use second-order Runge-Kutta time stepping, linear spatial reconstruction, and a Roe method Riemann solver. Parabolic terms due to viscosity are handled using a super-time-stepping technique, and we utilize the FARGO advection algorithm, which relaxes the restrictive Courant condition associated with the average orbital motion of the disk. We use a static polar (r, ϕ) grid with uniform Δr and $\Delta\phi$, with a canonical resolution of $N_r \times N_\phi = 512 \times 256$.

3.3 Canonical Run

The parameters of our canonical run (also referred to as "run 00" from here on) are as follows. The disk has an aspect ratio $h = 0.1$ and an active zone viscosity parameter $\alpha_0 = 0.1$. The DZ, in which the viscosity parameter is reduced by a factor $\epsilon_{\rm DZ} = 0.1$, extends from $r_{\rm IDZ} = 4.5$ to $r_{\rm ODZ} = 6.5$. The width of the viscosity transition at each DZ edge is equal to half of the local scale height, i.e., $\Delta r_{\rm IDZ} = h r_{\rm IDZ}/2 = 0.225$ and $\Delta r_{\rm ODZ} = h r_{\rm ODZ}/2 = 0.325$. The viscous timescale is $t_{\rm visc} \sim r^2/\nu \sim 160 (r/r_{\rm in})^{3/2} P_{\rm in}$, except in the DZ, where it is longer due to the reduction of α, and near the DZ edges, where it is shorter due to sharp viscosity gradients. Therefore, we run the simulation for $10^4 P_{\rm in}$, which captures several (~ 4) viscous timescales near $r_{\rm ODZ}$, and more than one viscous timescale at $r_{\rm out}$.

3.3.1 Evolution

The evolution of the canonical run is summarized in Fig. 3.1, which shows Σ at six representative times. Initially, the surface density in the DZ increases, as a result of the accretion of mass from the outer disk stalling in the DZ. Consequently, an overdense ring forms in the DZ ($t = 200$), and the disk remains axisymmetric as the amplitude of the bump grows. Eventually, it becomes large enough to trigger RWI. Once it saturates ($t = 270$), the most prominent density perturbations have azimuthal number $m = 4$, i.e., there are four vortices. These then merge into three ($t = 320$), and finally two vortices ($t = 400$). The remaining two vortices, positioned 180° apart in azimuth ($t = 500$), resist merging with one another, so the final morphology has an $m = 2$ symmetry, whose appearance changes only marginally over the remaining course of the simula-

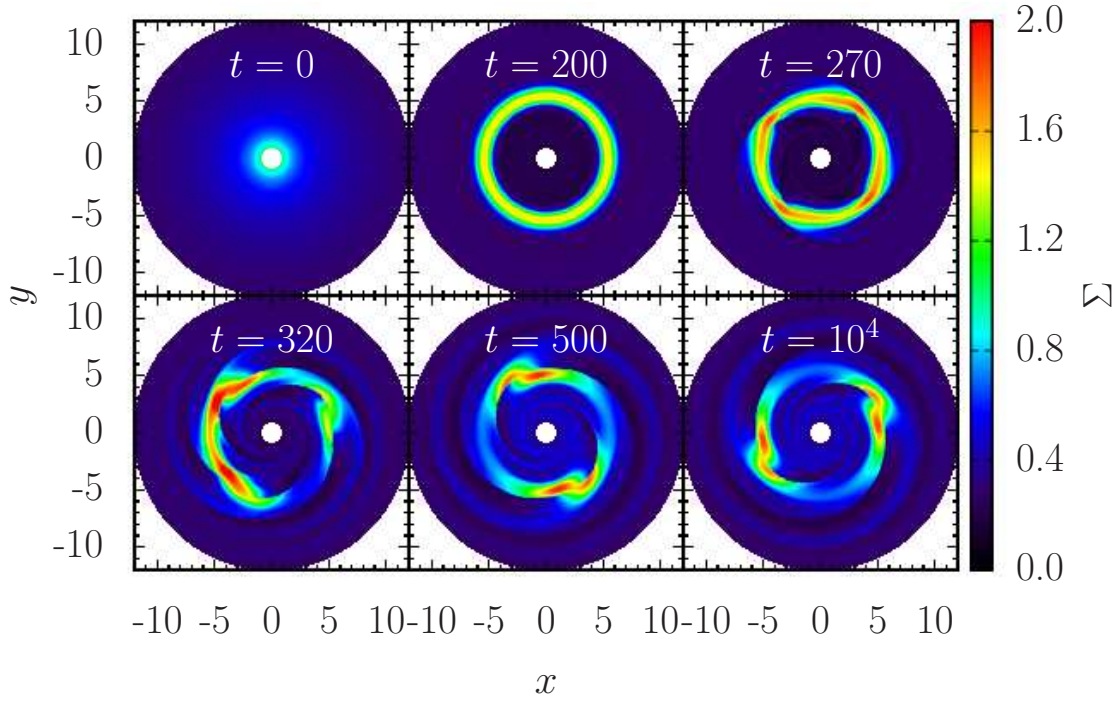

Figure 3.1: Surface density at several times during the evolution of the canonical run. From top left to bottom right, the panels represent: (i) the initial condition, (ii) the axisymmetric bump before becoming unstable, (iii) the saturated state of the RWI with a prominent $m = 4$ pattern, (iv) the nonlinear saturation of the RWI and merging of the four vortices into three, (v) further vortex merging, resulting in two vortices, (vi) the quasi-steady state, with two persistent vortices.

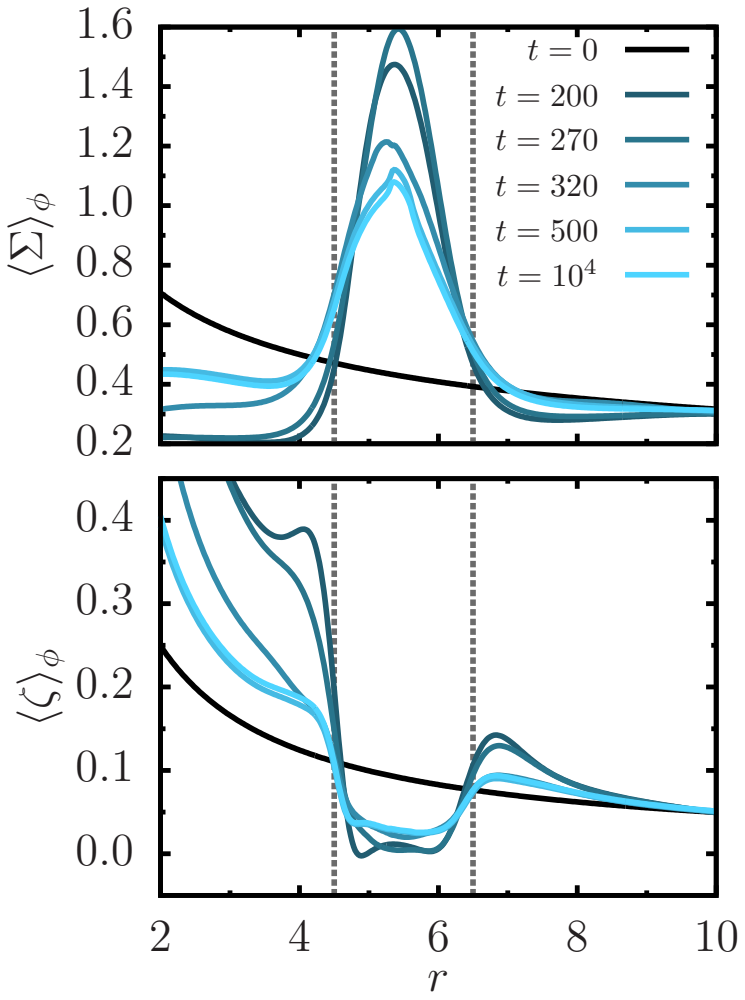

Figure 3.2: Azimuthally-averaged surface density and vortensity profiles at the same times as shown in Fig. 3.1.

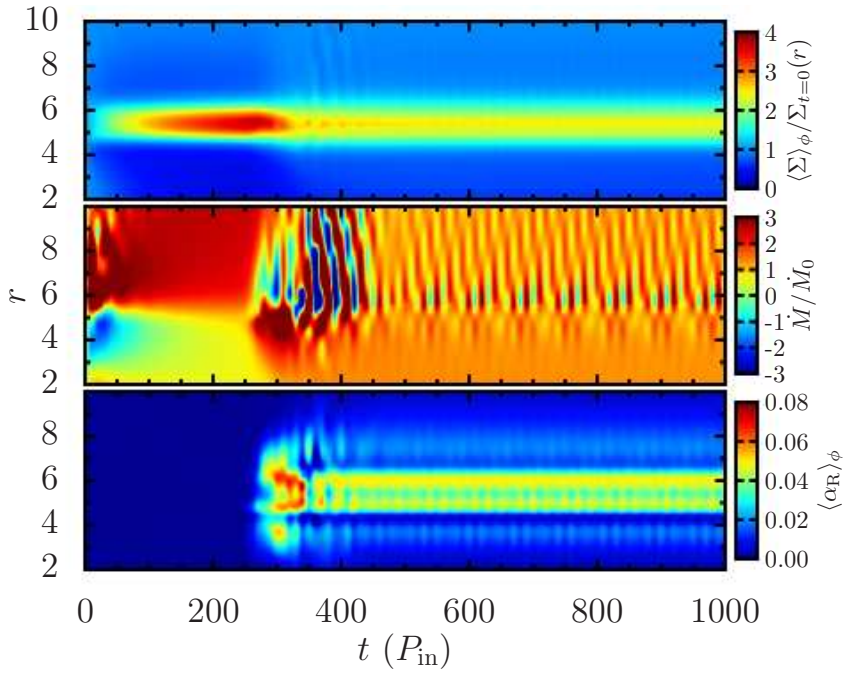

Figure 3.3: Evolution of the canonical run for $t < 1000\,P_{\rm in}$. From top to bottom, the panels show the azimuthally-averaged surface density $\langle\Sigma\rangle_\phi$ (normalized by the initial profile), the accretion rate $\dot{M}$, and the azimuthally-averaged Reynolds stress, $\langle\alpha_{\rm R}\rangle_\phi$, as functions of r and t.

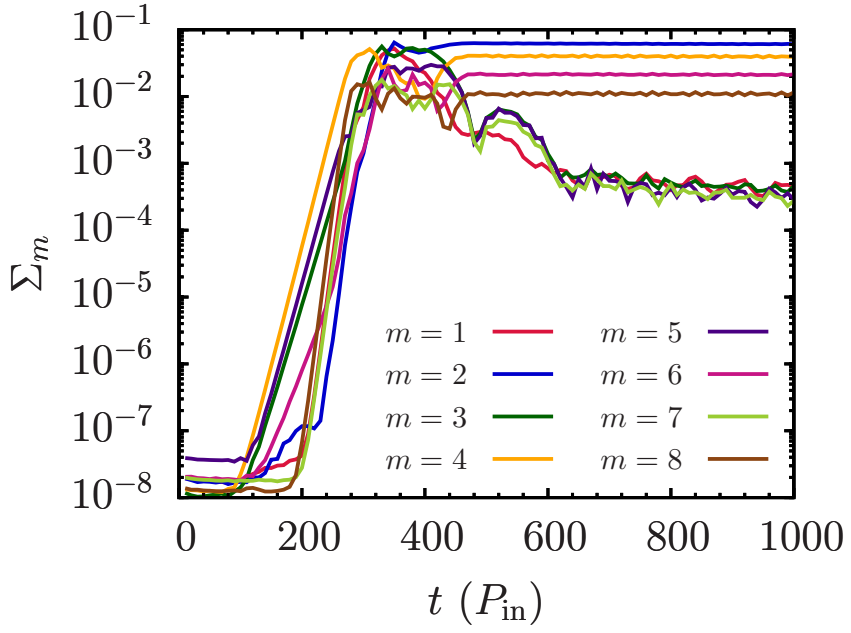

Figure 3.4: Evolution of Fourier components of surface density, Σ_m [Eq. (3.14)], in the canonical run.

tion ($t = 10^4$). Azimuthally-averaged profiles of Σ and vortensity $\zeta = |\nabla \times \mathbf{u}|/\Sigma$ at the same representative times are shown in Fig. 3.2. As the surface density in the DZ initially increases, forming a bump, RWI-unstable vortensity minima are produced. The peak of the density bump reaches a maximum that is ~ 3 times larger than the initial value before RWI is activated, which reduces the amplitude of the bump and smooths out the vortensity profile. By $t = 500$, both Σ and ζ have nearly reached their steady-state profiles, and only evolve a small amount subsequently. In the final state, the peak $\langle\Sigma\rangle_\phi$ corresponds to a factor of two enhancement relative to the initial value.

Further details of the evolution are illustrated in Fig. 3.3, which shows three azimuthally averaged quantities as functions of r and t: surface density Σ (normalized by the initial profile), accretion rate $\dot{M}$ (normalized by the supplied rate $\dot{M}_0$), and (dimensionless) Reynolds stress,

$$\langle\alpha_{\rm R}\rangle_\phi = \frac{\langle\Sigma\tilde{u}_r\tilde{u}_\phi\rangle_\phi}{\langle\Sigma\rangle_\phi c_{\rm s}^2}, \tag{3.13}$$

where $\tilde{u}_i = u_i - \langle u_i\rangle_\phi$. The development of a non-zero $\langle\alpha_{\rm R}\rangle_\phi$ at $t \approx 250$, which signifies enhanced angular momentum transport, is associated with the growth of the RWI, and is coincident with expulsion of mass from the DZ. This leads to a reduced peak $\langle\Sigma\rangle_\phi$, and a quasi-steady $\dot{M}$, which is modulated periodically, and with a large amplitude, by density waves, but has a steady time-averaged value close to $\dot{M}_0$ at all radii.

The growth, saturation and nonlinear evolution of the RWI are detailed in Fig. 3.4, which shows the integrated Fourier components of Σ,

$$\Sigma_m(t) = \frac{1}{2\pi(r_{\rm out} - r_{\rm in})}\int_{r_{\rm in}}^{r_{\rm out}}\int_0^{2\pi}\Sigma(r,\phi,t)\,{\rm e}^{-{\rm i}m\phi}{\rm d}\phi{\rm d}r, \tag{3.14}$$

as a function of time. Initially all components have approximately the same

small amplitude ($\sim 10^{-8}$), since the seed perturbations have no preferred azimuthal number. They begin to grow exponentially after about $100\,P_{\rm in}$, when the density bump becomes RWI-unstable. Note that although there are two vortensity minima at this point (see Fig. 3.2), they do not become unstable independent of each another. Instead, their proximity allows them to interact with one another, so that the entire DZ is effectively a single site for RWI to occur. During the growth phase, all of the Σ_m's have similar growth rates, but the $m = 4$ component grows fastest and is the first to saturate, with an amplitude of $\sim 10^{-2}$. By about $360\,P_{\rm in}$, the other components have saturated at comparable values. Energy is then transferred from high-m to low-m modes through vortex merging (see Fig. 3.1). In the final state ($t > 600$), Σ_2 is largest, followed by Σ_4, Σ_6, and Σ_8, each of which is smaller than the previous by a factor of a few. The amplitudes of odd-m components are several orders of magnitude smaller. This configuration can be interpreted as a nonlinear $m = 2$ mode.

3.3.2 Quasi-Steady State

After about $600\,P_{\rm in}$, the disk reaches a global quasi-steady state. In this stage, the azimuthally-averaged profiles of Σ and $\alpha_{\rm R}$ remain steady, as illustrated by Fig. 3.3. The accretion rate varies periodically in much of the disk due to waves launched from the DZ, but is steady in a time-averaged sense. The waves are damped in the inner disk ($r \lesssim 4$), resulting in almost steady accretion through the inner boundary. Further, the azimuthal structure of the disk (i.e., the amplitude of density perturbations proportional to $e^{im\phi}$) does not evolve further (see Fig. 3.4). This azimuthal structure is of interest because it is dominated by an $m = 2$ pattern. This differs from the morphology resulting from an isolated den-

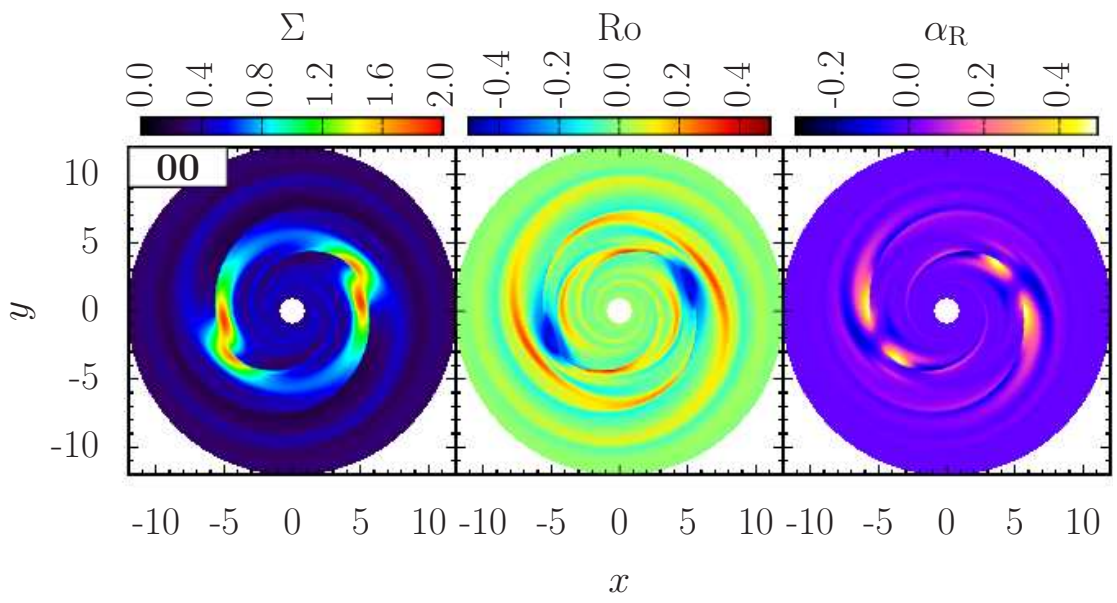

Figure 3.5: Snapshot of surface density Σ, Rossby number Ro, and Reynolds stress α_R in the quasi-steady state of the canonical run.

sity bump in an inviscid disk (e.g., Meheut et al. 2012b), in which all of the initial RWI vortices merge into a single vortex, resulting in a global $m = 1$ symmetry. The two vortices resist merging for the entire $10^4\ P_{in}$ duration of the simulation, which is much longer than the viscous timescale at the outer edge of the DZ. This morphology also appears in several other runs in our parameter study (see Section 3.4).

The global structure of the disk in the quasi-steady state is illustrated in Fig. 3.5, which shows snapshots of surface density, Rossby number, and local (dimensionless) Reynolds stress $\alpha_R = \tilde{u}_r \tilde{u}_\phi / c_s^2$. Here the Rossby number is defined

$$\mathrm{Ro} = \frac{\tilde{\omega}}{2\Omega_K}, \tag{3.15}$$

where $\tilde{\omega} = \left[\nabla \times \left(\boldsymbol{u} - r\Omega_K \hat{\boldsymbol{\phi}}\right)\right]_z$ is the residual (total minus Keplerian) vorticity. The nonlinear $m = 2$ morphology is apparent in the snapshots of all three quantities. The surface density features associated with this azimuthal symmetry

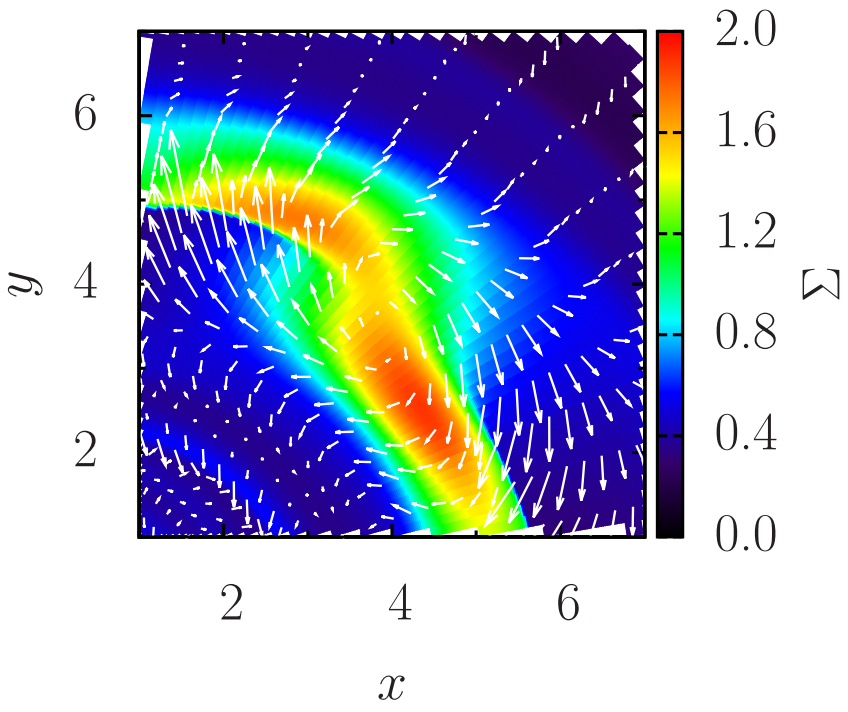

Figure 3.6: Zoom-in of one of the vortices in the canonical run (there is another, nearly identical vortex on the opposite side of the disk), showing surface density (colors) and velocity (with the Keplerian velocity subtracted; arrows). The longest arrows represent velocities approximately 1.8 times the local sound speed. The direction of the bulk disk rotation is anti-clockwise, so the clockwise motion of the velocity perturbations indicates that the vortex is anticylonic.

are coincident with regions of approximately constant Ro, which is characteristic of vortices. They have Ro ≈ -0.5, or $\tilde{\omega} \approx -\Omega_K$, which means that they approximately rotate with the local shear (for a Keplerian flow, the local shear is $-3\Omega_K/2$), and thus their survival is not threatened by the shear. The center of the vortices have small Reynolds stresses, while large positive stress is produced at their edges. The angular momentum transport associated with the vortices and the associated spiral density waves outside of the DZ is described by the azimuthally-averaged Reynolds stress (see Fig. 3.3), which can be interpreted as an effective viscosity parameter. It is peaked at about 0.05 (near the outer DZ edge), but remains larger than 0.03 even in the center of the DZ. Since this is larger than the intrinsic DZ viscosity, $\epsilon_{DZ}\alpha_0 = 0.01$, it plays a significant role in maintaing steady accretion through the DZ.

Figure 3.6 gives a detailed view of one of the two persistent vortices in the quasi-steady state. It consists of two dense blobs which rotate anticyclonically about a slightly less dense core (although the core has a surface density several times larger than the surrounding disk). The shape of the vortex is notable, as it differs significantly from that of a typical RWI vortex, which consists of a smoother surface density profile with only a single maximum. Although the overdense feature is large, with a radial extent similar to that of the DZ itself, and spanning nearly 90° in azimuth, the vortex proper, i.e., the region of negative vorticity (see Fig. 3.5) is smaller, with $\Delta r \approx 1.5$ and $r\Delta\phi \approx 4.0$. The largest velocity perturbations associated with it have an amplitude $v_{max} \approx 1.8\, c_s$ (with c_s evaluated in the middle of the DZ). The radial size of the vortex is roughly consistent with $|v_{max}/\tilde{\omega}| \approx 1.8H$ (where $\tilde{\omega} \approx -\Omega_K$), the distance over which it can remain coherent given its rotation frequency and velocity. While the average surface density in the DZ is about two times larger than it would be in a con-

stant α disk, the maximum surface density enhancement (in the centers of the blobs) is larger by a further factor of two.

3.3.3 Periodicity

A notable feature of the canonical run (as well as other runs in our parameter study) is that the velocity field of the disk is globally periodic, despite the presence of highly nonlinear waves (see Fig. 3.7). Figure 3.9 (top left panel) shows the power spectrum of radial velocity u_r, i.e., the square of its temporal Fourier transform,

$$\mathcal{F}\left[u_r\right](r,\omega) = \int_{t_1}^{t_2} u_r\left(r,0,t\right) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{dt}, \tag{3.16}$$

where t_1 and t_2 are in the quasi-steady state of the disk with $t_2 - t_1 = 100\,P_{\mathrm{in}}$. Power is concentrated in a global mode with frequency $\omega = 0.159\,\Omega_{\mathrm{in}}$ (and harmonics of this fundamental frequency). Delineating the power spectrum by azimuthal number reveals that it is an $m = 2$ mode, so this frequency is twice the pattern frequency of the mode. The pattern frequency, $\omega_{\mathrm{P}} = 0.079\,\Omega_{\mathrm{in}}$ is very close to the Keplerian frequency at the peak of the azimuthally-averaged surface density profile (located at $r_{\mathrm{peak}} = 5.38$), $\Omega_{\mathrm{K,peak}} = 0.080\,\Omega_{\mathrm{in}}$. This is characteristic of RWI modes, whose frequencies are m-times the corotation frequency of the unstable bump, but it is remarkable that this coherent global frequency persists in its nonlinear stage.

The global oscillations are quasi-periodic, and the power is spread around the peak frequency with a full-width half-maximum of $\Delta\omega/\omega = 8.5\%$. Thus, the velocity fluctuations at any (Eulerian) point, sampled one period apart in time, are not exactly equal to one another. Figure 3.7 depicts the velocities u_r and u_ϕ

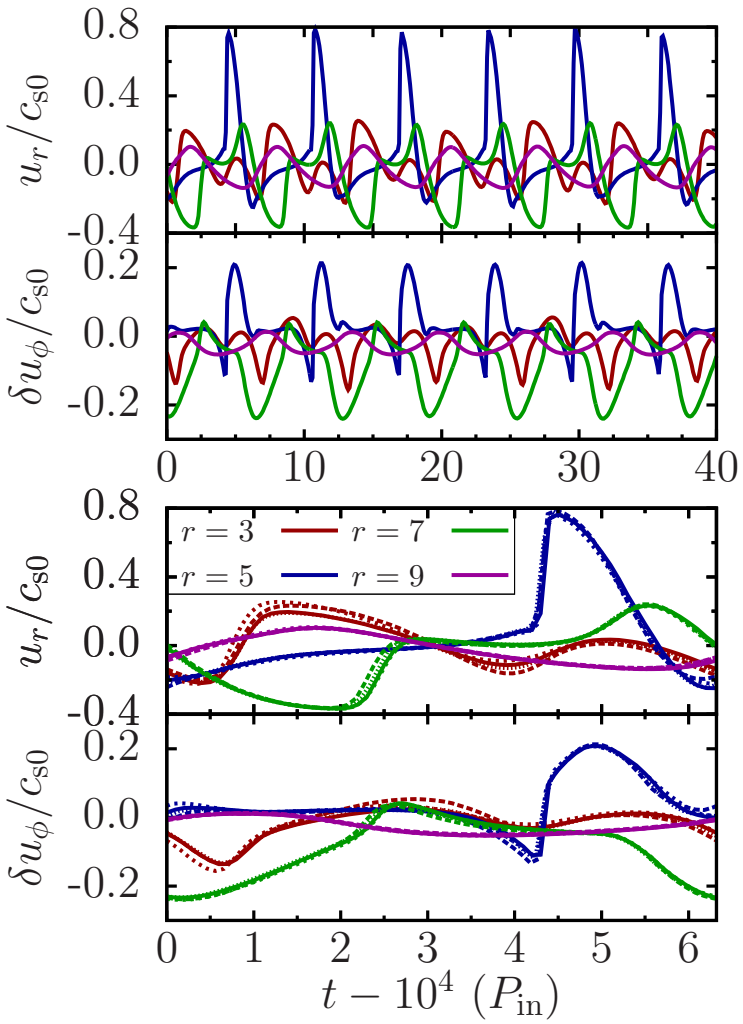

Figure 3.7: Radial and azimuthal (minus Keplerian) velocities at several locations, specified by $(r, \phi) = (r, 0)$, as a function of time during the quasi-steady state of the canonical run. The top panel shows several oscillation periods, and the bottom panel shows four of them (represented by different line types, e.g., solid, dashed, etc.) folded with a period of 6.31 P_{in}.

at several radii (evaluated at $\phi = 0$) as a function of time for several oscillation periods after the main $10^4\, P_{\rm in}$ run. The same periodicity is seen at all locations, since they are the result of a global mode. The lower panel gives a closer look at the quasi-periodicity by folding over four periods (with a period extracted from the power spectrum), demonstrating the slight variation between successive oscillation periods.

3.3.4 Resolution and Convergence

We performed the canonical run with double the number of radial and azimuthal grid points ($N_r \times N_\phi = 1024 \times 512$; we denote this run "00_res") to determine how our our results depend on resolution. No qualitative differences are observed in the disk evolution (e.g., development of bump, growth of RWI), or the morphological appearance in the quasi-steady state ($m = 2$ symmetry and two-lobed vortex shape). The key numerical quantities characterizing the quasi-steady are given in Table 3.1, alongside those of the standard canonical run. They differ by less than 2%, which indicates that they are converged with respect to resolution in the standard run. We also performed the canonical run with a steeper initial surface density profile ($\Sigma \propto r^{-1}$), and found that it does not impact the final disk configuration, only the transient initial growth of the bump and the RWI. Therefore, the quasi-steady state that is eventually reached is not sensitive to the initial conditions, and only depends on the DZ geometry.

Run	h	α_0	ϵ_{DZ}	r_{IDZ}, r_{ODZ}	$\Delta r_{IDZ/ODZ}$	M_{DZ}/M_{nodz}	$\Sigma_{max}/\Sigma_{0,dz}$	$\langle\langle\alpha_R\rangle\rangle_{\phi,t,max}$	$\langle\langle\langle\alpha_R\rangle\rangle_{\phi,t}\rangle_{DZ}$	m
00	0.1	0.1	0.1	4.5, 6.5	$H/2$	1.96	4.34	4.89×10^{-2}	3.97×10^{-2}	2
00_res	0.1	0.1	0.1	4.5, 6.5	$H/2$	1.93	4.26	4.89×10^{-2}	4.03×10^{-2}	2
R1	0.1	0.1	0.1	4.0, 7.0	$H/2$	4.71	8.71	1.89×10^{-2}	8.62×10^{-3}	2/2
R2	0.1	0.1	0.1	3.5, 7.5	$H/2$	5.90	11.10	1.34×10^{-2}	4.10×10^{-3}	1/2
R3	0.1	0.1	0.1	2.5, 8.5	$H/2$	6.50	11.38	8.92×10^{-3}	2.19×10^{-3}	1/1
DR1	0.1	0.1	0.1	4.5, 6.5	H	2.09	4.11	2.58×10^{-2}	2.33×10^{-2}	2
DR2	0.1	0.1	0.1	4.5, 6.5	$3H/2$	2.41	3.66	1.53×10^{-3}	5.85×10^{-4}	2
E1	0.1	0.1	0.01	4.5, 6.5	$H/2$	2.05	5.07	6.00×10^{-2}	5.03×10^{-2}	2
C1	0.075	0.1	0.1	4.5, 6.5	$H/2$	3.40	14.20	1.71×10^{-2}	1.27×10^{-2}	1
C2	0.05	0.1	0.1	4.5, 6.5	$H/2$	6.21	11.31	1.05×10^{-2}	3.52×10^{-3}	3/2
A1	0.1	0.05	0.1	4.5, 6.5	$H/2$	1.65	3.47	3.16×10^{-2}	2.63×10^{-2}	2
A2	0.1	0.025	0.1	4.5, 6.5	$H/2$	1.81	3.69	1.42×10^{-2}	1.14×10^{-2}	1

Table 3.1: Parameters and main quantitative results of our simulations. The first column gives the alphanumeric label of each run. The next five columns give the parameters of the runs: disk aspect ratio $h = H/r$, active zone viscosity parameter α_0, DZ viscosity reduction factor ϵ_{DZ}, DZ edges r_{IDZ} and r_{ODZ}, and viscosity transition widths Δr_{IDZ} and r_{ODZ} (see Section 3.2). The next column gives the mass in the DZ, M_{DZ}, in the quasi-steady state, normalized by the mass at $t = 0$. The next column is the maximum surface density Σ_{max}, relative to the average density in the middle of the DZ for a constant α disk. The next two columns give the maximum of the azimuthally-averaged and time-averaged (over 100 P_{in} in the quasi-steady state) dimensionless Reynolds stress, $\langle\langle\alpha_R\rangle\rangle_{\phi,t,max}$, and its average in the DZ, $\langle\langle\langle\alpha_R\rangle\rangle_{\phi,t}\rangle_{DZ}$. The last column describes the morphology of the quasi-steady state in terms of the dominant azimuthal mode number m. A single value in this column indicates the azimuthal number of a single global mode, which is coherent across the entire DZ. Two values separated by a slash indicate the m's of two modes, localized to the outer (first value) and inner (second value) DZ edges.

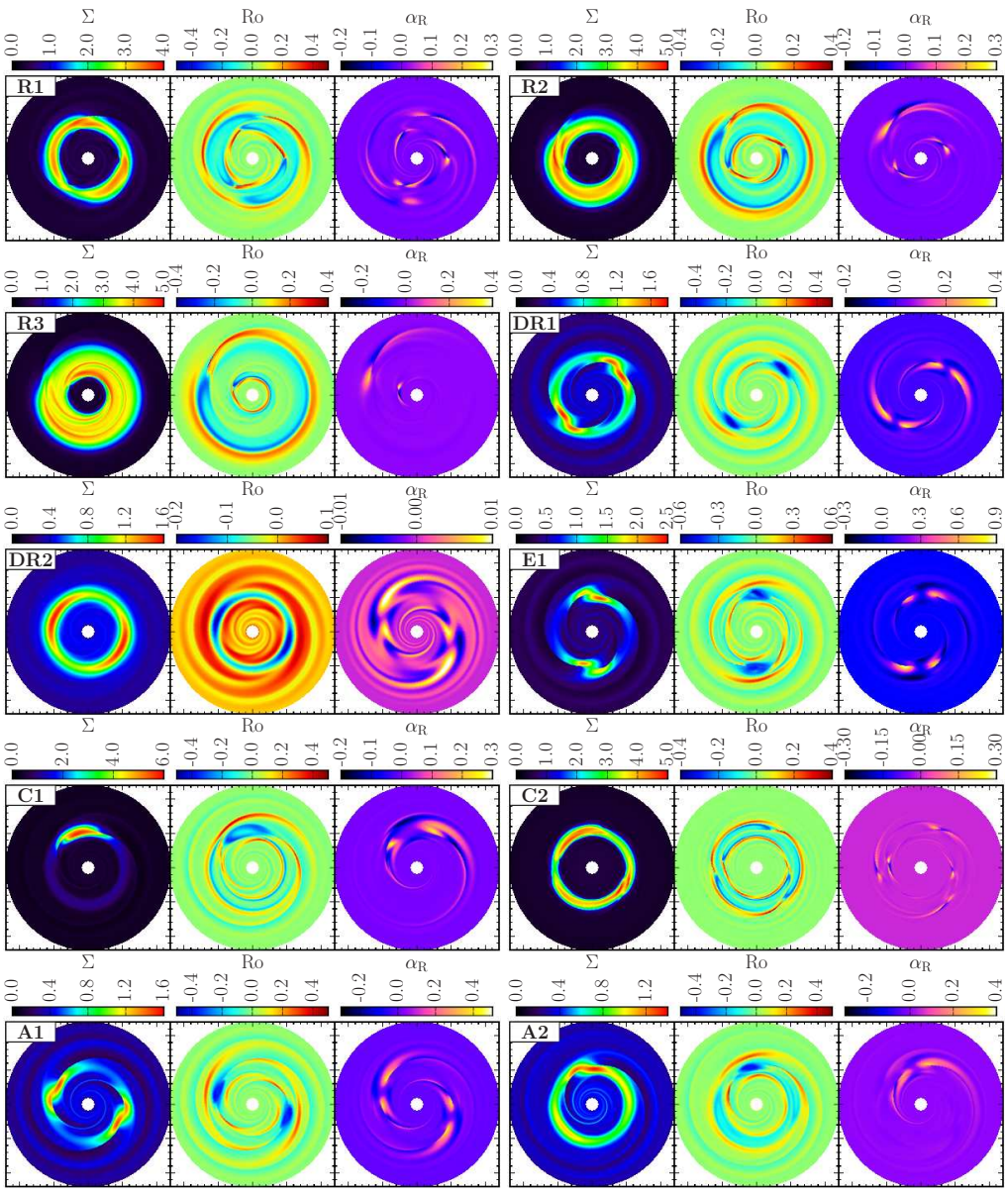

Figure 3.8: Snapshots of surface density, Rossby number, and Reynolds stress, in the quasi-steady state (as in Fig. 3.5), for all runs except the canonical run. The scale of the x and y axes are the same as in Fig. 3.5, but the color scales for the three plotted quantities differ between panels.

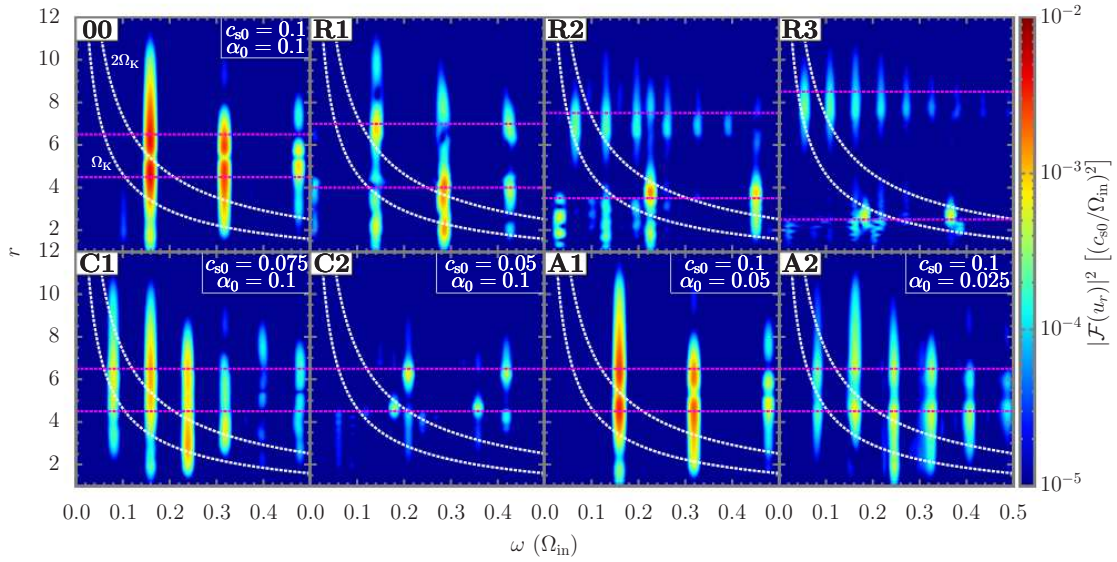

Figure 3.9: Power spectrum of u_r as a function of r for eight different runs. The extent of the DZ, which is varied in the top four panels, is indicated by the horizontal dashed lines. The dashed curves labeled Ω_K and $2\Omega_K$ correspond to the Keplerian orbital frequency and twice the Keplerian orbital frequency.

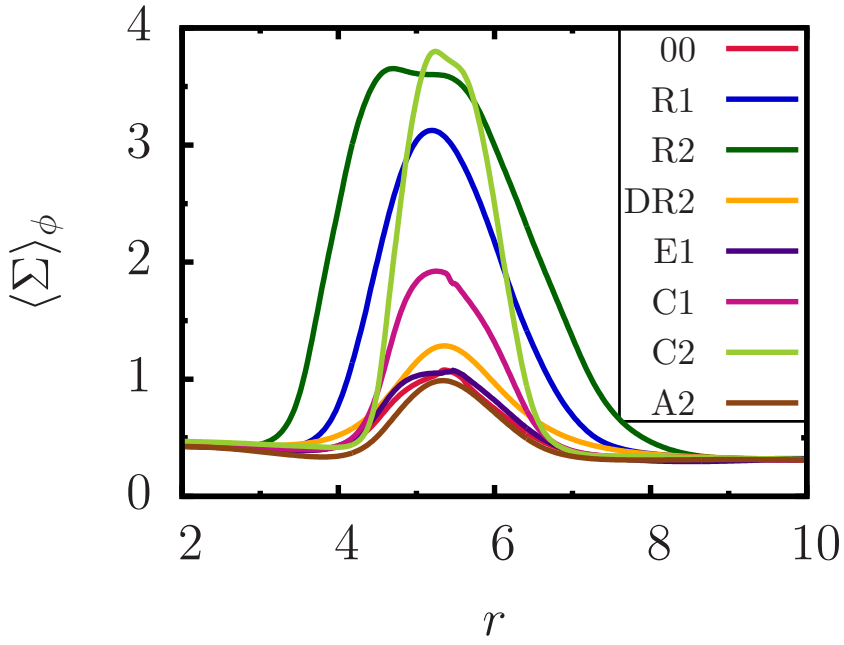

Figure 3.10: Azimuthally-averaged surface density profiles in the quasi-steady state for eight of the eleven runs.

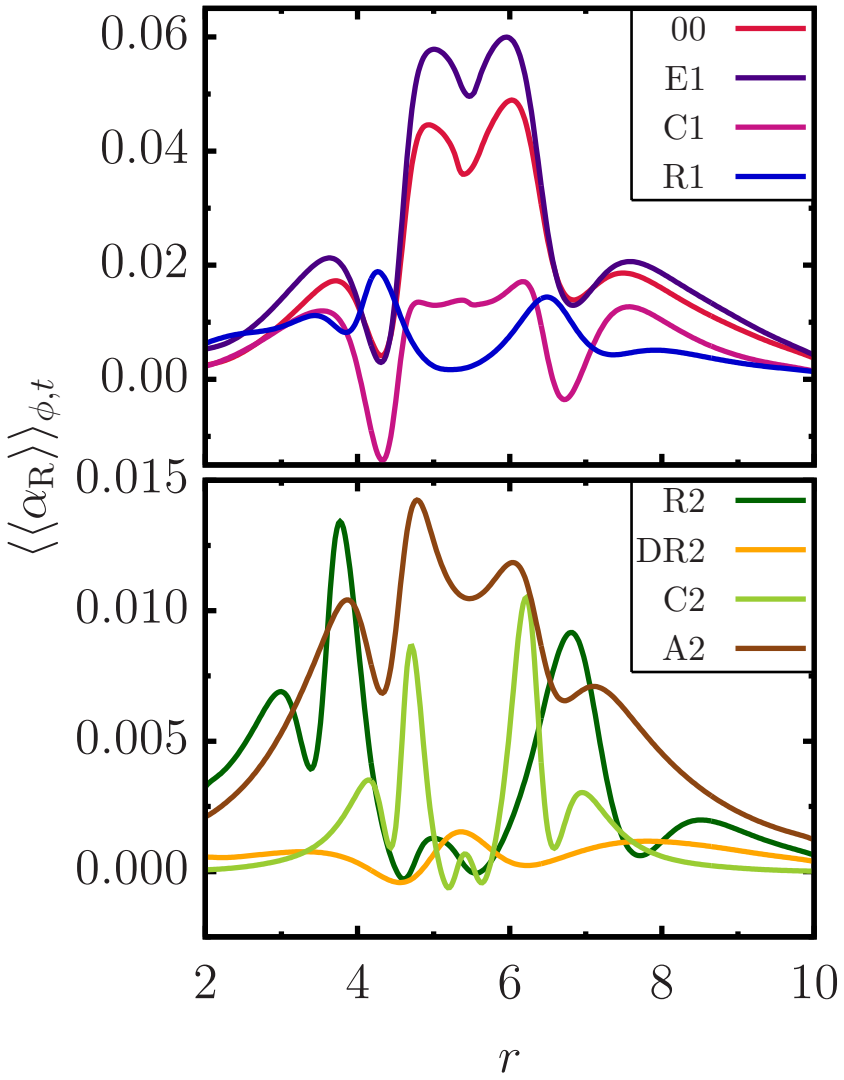

Figure 3.11: Double-averaged Reynolds stress (azimuthally-averaged and time-averaged for $100\,P_{\rm in}$) for eight of the eleven runs. Note that the range of the y-axis is different in the top and bottom panels. The peak and DZ-averaged values are listed in Table 3.1.

3.4 Parameter Study

To test the dependence of the angular momentum transport and morphology on the DZ parameters, we ran a suite of simulations, each differing from the canonical run by the value of a single parameter or pair of related parameters. The parameters for each run are listed in Table 3.1. Each run is labeled with letters, indicating which parameter is being varied (R for DZ width, DR for viscosity transition width, E for viscosity reduction factor, C for sound speed, and A for active zone viscosity), and a number, which indexes multiple runs with different parameter values. Most runs, like the canonical run 00, are evolved for $10^4\,P_{\rm in}$, but runs R3, A1 and A2 are evolved for $1.5\times10^4\,P_{\rm in}$, and C2 for $2.5\times10^4\,P_{\rm in}$. These durations are long enough to allow the disk to reach a quasi-steady state, in which there are no longer appreciable changes in the Fourier components of Σ, or its azimuthally-averaged profile. While the former typically happens by about $1000\,P_{\rm in}$, shortly after the RWI has saturated, the latter does not happen until much later, requiring viscous timescales to reach.

Throughout this section we refer to several figures which illustrate the main results of our parameter study simulations. Figure 3.8 shows a snapshot of the surface density, Rossby number and Reynolds stress at the end of each simulation (as in Fig. 3.5 for run 00). These illustrate the morphology, surface density contrast, presence or absence of vortices, and level of angular momentum transport in the quasi-steady state. Azimuthally-averaged surface density profiles for eight of the eleven simulations are shown in Fig. 3.10, and double-averaged (i.e., averaged over azimuth and time) dimensionless Reynolds stress $\langle\langle\alpha_{\rm R}\rangle\rangle_{\phi,t} = \langle\langle\Sigma\tilde{u}_r\tilde{u}_\phi\rangle\rangle_{\phi,t}/(\langle\langle\Sigma\rangle\rangle_{\phi,t}c_{\rm s}^2)$ profiles in Fig. 3.11. Together these indicate the efficiency of RWI-driven angular momentum transport (large Reynolds stress

and low DZ surface density indicate efficient transport). Figure 3.9 illustrates the power spectrum (temporal Fourier transform squared) of u_r taken over 100 orbits in the quasi-steady state for eight of the simulations. This demonstrates the variety of oscillations present in the disk, the spatial extent of their coherence, and how they are affected by the width of the DZ and by viscosity (as described in Sections 3.4.1 and 3.4.5).

Table 3.1 summarizes several key quantities characterizing the quasi-steady state for each simulation. The mass contained in the DZ, $M_{\rm DZ}$, is expressed in terms of the mass in the DZ at $t = 0$,

$$M_{\rm nodz} = \int_0^{2\pi} \int_{r_{\rm IDZ}}^{r_{\rm ODZ}} r^{-\frac{1}{2}} r \mathrm{d}r\mathrm{d}\phi, \tag{3.17}$$

which is equal to the mass between $r_{\rm IDZ}$ and $r_{\rm ODZ}$ for a steady-state axisymmetric disk with $\epsilon_{\rm DZ} = 1$, i.e., for a constant-α disk with no DZ (hence the subscript "nodz"). The maximum surface density $\Sigma_{\rm max}$ is given in terms of the surface density in the middle of the DZ at $t = 0$, $\Sigma_{0,\rm DZ}$ (equal to $1/\sqrt{5.5} = 0.43$ in all runs). This indicates the largest factor by which the surface density is enhanced relative to the ambient density slightly outside of the DZ. The maximum and DZ-averaged values of the double-averaged Reynolds stress, $\langle\langle\alpha_{\rm R}\rangle\rangle_{\phi,t}$, computed from the profiles in Fig. 3.11, are also listed in Table 3.1.

3.4.1 Dead Zone Width

The radial width of the DZ is varied in runs R1, R2 and R3, by decreasing $r_{\rm IDZ}$ and increasing $r_{\rm ODZ}$. The width of the viscosity transitions, $\Delta r_{\rm IDZ/ODZ}$, are scaled with the local scale height as in the canonical run ($\Delta r_{\rm IDZ/ODZ} = H/2$), and thus differ in absolute size between these runs. It is useful to define the dimensionless

DZ width

$$\Delta_{\rm DZ} = \frac{2\,(r_{\rm ODZ} - r_{\rm IDZ})}{r_{\rm IDZ} + r_{\rm ODZ}}, \tag{3.18}$$

i.e., the width of the DZ divided by the the radial coordinate of its center. The canonical run (and all others besides the "R" runs) has $\Delta_{\rm DZ} = 0.36$, while runs R1, R2 and R3 have $\Delta_{\rm DZ} = 0.55, 0.73$ and 1.09. These four runs demonstrate the features of "narrow" and "wide" DZs, as well as the intermediate behavior between these two regimes.

The main effect of varying the width of the DZ is shown in the top four panels of Fig. 3.9. Here the power spectrum of radial velocity is shown for runs 00 and R1-R3, with increasing $\Delta_{\rm DZ}$ from left to right. For the smallest DZ width (as detailed in Section 3.3), there is a global, coherent oscillation mode with a frequency equal to twice (i.e., an $m = 2$ mode) the Keplerian orbital frequency (evaluated at the DZ center) spanning the entire DZ, as well as its higher frequency harmonics. This represents the typical behavior of a "narrow" DZ. For the largest DZ width (run R3), there is a distinct mode present near each DZ edge, each corresponding to the Keplerian frequency at a radius close to the edge (i.e., they are $m = 1$ modes), along with harmonics. Each mode is coherent only near its respective DZ edge, and does not effectively interact with the other edge. This represents a typical "wide" DZ. Intermediate separations (runs R1 and R2; the top middle panels of Fig. 3.9) lead to more complicated behaviors in between those of narrow and wide DZs. In these cases, there are two distinct modes, each with a fundamental frequency associated with one DZ edge, but not necessarily with the same m. For example, R2 has an $m = 2$ mode near the inner DZ edge [i.e., its frequency is approximately $2\Omega_{\rm K}(r_{\rm IDZ})$] and an $m = 1$ mode near the outer edge. There are varying degrees of coherence between the edge modes, but they are distinctly less coherent than for the narrow DZ case. The

transition between the narrow and wide DZ regimes occurs at $\Delta_{DZ} \approx 0.7$.

As the DZ width increases, the regions of negative vorticity become highly azimuthally elongated, and are not coincident with distinct overdensities (see Fig. 3.8 for runs R1-R3 and Fig. 3.5 for run 00). The effective Reynolds stress also decreases (see Table 3.1), with only run R1 having $\langle \alpha_R \rangle_{DZ} \sim 10^{-2}$ (recall that 10^{-2} is the residual intrinsic DZ viscosity). Consequently, the mass in the DZ for runs R1-R3 is two to three times larger than in the canonical run.

3.4.2 Viscosity Transition Width

We explore the effects of changing the viscosity transition widths in runs DR1 and DR2, in which Δr_{IDZ} and Δr_{ODZ} are two and three times larger than in run 00. In its quasi-steady state, DR1 is very similar to the canonical run, in terms of morphology, DZ mass (slightly larger than the canonical run), and peak/DZ-averaged Reynolds stress (slightly smaller than the canonical run). Run DR2 is dramatically different, with the Reynolds stress reduced by two orders of magnitude, and a morphology which resembles that of the linear-phase RWI (e.g., the third panel of Fig. 3.1), rather than having the strong nonlinear features of the canonical run. The evolution of the Σ_m's reveals that their exponential growth is halted at a smaller amplitude than in the canonical run, and subsequently maintained at that amplitude, suggesting that the RWI is partially suppressed by viscosity.

This behavior can be understood as follows. The broader viscosity transitions reduce the sharpness of the density bump which develops in the DZ. This has two consequences on the linear RWI: the intrinsic growth rate is reduced,

and the viscous damping rate is also reduced (i.e., the viscous diffusion time across the bump increases). As shown in Appendix B, competition between these two effects results in the existence of a critical α above which the RWI is suppressed. We estimate that for a bump of width of H, $\alpha_{\rm crit} \approx 0.04$. Thus, in run DR1, the RWI is not strongly affected by viscosity, since $\alpha_{\rm DZ} = 0.01 < \alpha_{\rm crit}$. An estimate of $\alpha_{\rm crit}$ relevant to run DR2 suggests that the growth of the RWI is significantly reduced by viscosity. Strictly, our viscous damping criterion only applies to the linear RWI, but we expect the nonlinear evolution to be affected to a similar degree. This explains the qualitatively different outcome of run DR2 compared to DR1. We conclude that efficient revival of the DZ by the RWI requires $\Delta r_{\rm IDZ}, \Delta r_{\rm ODZ} \lesssim H$ (in agreement with previous studies, e.g., Lyra et al. 2009b; Regály et al. 2012).

3.4.3 Viscosity Reduction Factor

In run E1, we set $\epsilon_{\rm DZ} = 0.01$, ten times smaller than in the canonical run. In an axisymmetric, purely viscously evolving disk, this would result in a surface density enhancement of $\sim \epsilon_{\rm DZ}^{-1} = 100$ (compared to a constant α disk) in the DZ. However, once a quasi-steady state has been reached, the mass accumulated in the DZ is only about 5% larger than in the canonical run. All other features of this run are also very similar, such as the azimuthal symmetry and vortex shape (see Fig. 3.8). The main difference is the strength of angular momentum transport; the peak and DZ-averaged Reynolds stresses are $\sim 25\%$ larger. This may be due to the fact that the reduced DZ viscosity leads to less viscous damping of the vortex and density wave motions, allowing them to be more vigorous. This behavior is in qualitative agreement with our analysis of the effect of vis-

cosity in the linear regime (see Appendix B). Overall, compared to other DZ parameters, the value of ϵ_{DZ} has a relatively minor effect on the properties of the quasi-steady state of the disk.

3.4.4 Sound Speed

The effects of lowering the sound speed, by means of lowering the aspect ratio h, are illustrated in Fig. 3.9. For $h = 0.075$ (run C1), the results are similar to run 00, where a coherent global oscillation develops in the DZ. However, it is dominated by an $m = 1$ mode (i.e., $\omega \approx \Omega_K$) rather than an $m = 2$ mode. This may be related to the role of viscosity, since $\nu \propto \alpha h^2$, and a similar effect is seen when α_0 is reduced (see Section 3.4.5). The Reynolds stress is reduced by a factor of three and the mass in the DZ increases by about 75%.

When h is further reduced to 0.05 (run C2), the behavior becomes similar to the wide DZ case described in Section 3.4.1. Oscillations localized at each DZ edge are present, with azimuthal numbers greater than unity, and with relatively low amplitude in the power spectrum. The Reynolds stress becomes small, about an order of magnitude smaller than in run 00, and the mass in the DZ becomes large (three times larger than run 00) to compensate for the low level of stress. In this case, the wide DZ behavior occurs because the sound speed is too small to allow the entire DZ to oscillate coherently. In order for both DZ edges to interact with one another, the sound crossing time of the DZ, $t_{cross} \sim (r_{ODZ} - r_{IDZ})/\langle c_s \rangle_{DZ}$, should be less than the period of a density wave produced by the RWI, which is approximately the local orbital period, i.e., the ratio $t_{cross}/P_{orb,DZ} \sim \Delta_{DZ}/(2\pi h)$ should be less unity. This criterion is satisfied in run C1,

but not in run C2.

3.4.5 Active Zone Viscosity

Run A1, in which α_0 is reduced to half of its canonical run value, produces a quasi-steady state similar to that of run 00. The morphology is dominated by the same $m = 2$ mode and the shape of the vortex is the same. The DZ mass and maximum surface density are about 20% smaller, and the Reynolds stress (either the peak or DZ-averaged value) is 35% smaller. This suggests that the disk self-regulates the Reynolds stress produced in the DZ, in order to achieve a similar level of angular momentum transport as in the active zone.

Further reducing α_0 in run A2 leads to many similar properties in the quasi-steady state. There is a small increase in $M_{\rm DZ}$ and $\Sigma_{\rm max}$, and a corresponding reduction in $\alpha_{\rm R}$. However, there is a major change in the morphology of the disk. The most prominent azimuthal mode number becomes $m = 1$ rather than $m = 2$. A similar morphological transition occurs when the sound speed is reduced (see Section 3.4.4). In both cases, this change is associated with increasing the Reynolds number, $\mathrm{Re} \sim r^2\Omega_{\rm K}/\nu \sim \alpha_0^{-1}h^{-2}$ (in the active zone; note that the Reynolds number in the DZ does not affect the disk morphology, see Section 3.4.3), which describes the importance of inertial forces relative to viscous forces. Based on the values of the Reynolds number in runs 00, C1, A1, and A2, we conclude that a narrow DZ ($\Delta_{\rm DZ} \lesssim 0.7$) with $\mathrm{Re} \lesssim 1800 - 2000$ leads to an $m = 2$ morphology, otherwise an $m = 1$ morphology is produced.

The physical origin of the $m = 1$ to $m = 2$ morphological change is not clear. We speculate that it is related to the self-regulation of the stress produced in the

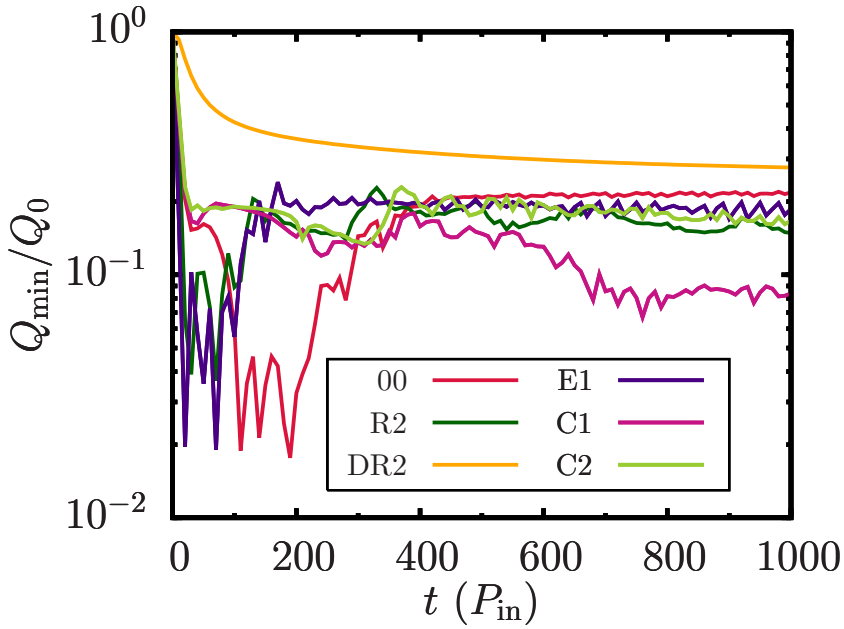

Figure 3.12: Minimum Toomre parameter in the DZ, $Q_{\rm min}$, as a function of time, normalized by Q_0 [see Eq. (3.20)], for $t < 1000\,P_{\rm in}$ in several runs.

DZ in order to match the viscous stress in the active zone. For large viscosities, the stress associated with the $m = 1$ mode is insufficient to match the angular momentum transport in the active zone, and so its amplitude is reduced relative to the $m = 2$ mode, which can sustain stronger transport.

3.4.6 Gravitational Stability

Self-gravity is not included in our simulations. To evaluate its possible importance, we calculate the local value of Toomre Q parameter,

$$Q = \frac{\kappa c_{\rm s}}{\pi G \Sigma}. \tag{3.19}$$

Since in our simulations, Σ is scale free, we scale Q with respect to a reference value. For an axisymmetric α-disk in steady state, the surface density is related

to the accretion rate by $\dot{M} = 3\pi\nu\Sigma$, and the Toomre parameter at radius $r_{\rm ODZ}$ is given by

$$Q_0 = 590\left(\frac{\alpha}{0.01}\right)\left(\frac{h}{0.1}\right)^3\left(\frac{r_{\rm ODZ}}{10\,{\rm AU}}\right)^{-3/2} \times\left(\frac{M}{1\,M_\odot}\right)^{3/2}\left(\frac{\dot{M}}{10^{-8}\,M_\odot/{\rm yr}}\right)^{-1}, \tag{3.20}$$

which we take as our reference value. We have adopted some fiducial parameters, including a typical accretion rate for protoplanetary disks.

As mass accumulates in the DZ, the Q value in the DZ becomes smaller than Q_0. Figure 3.12 shows the minimum local value of Q in the DZ, $Q_{\rm min}$, normalized by Q_0, as a function of time (for $t < 1000\,P_{\rm in}$) for several runs. We see that before the RWI develops, the ratio $Q_{\rm min}/Q_0$ can be as small as 0.02 (e.g., run 00). Once the quasi-steady state is reached, $Q_{\rm min}/Q_0$ settles to a modest value between 0.08 and 0.3, depending on the parameters of the simulation. The largest reduction ($Q_{\rm min}/Q_0 \approx 0.08$) occurs for run C1 (reduced sound speed), which is also the run for which $\Sigma_{\rm max}/\Sigma_{0,\rm DZ}$ is largest (see Table 3.1).

We conclude that, because of the angular momentum transport associated with the nonlinear RWI, the Toomre Q parameter in the DZ can be reduced from the fiducial value Q_0 by at most a factor of 12, for a wide range of DZ parameters. Therefore, Eq. (3.20) can be used to estimate the parameters of the steady-state disk (just outside $r_{\rm ODZ}$) for which the DZ will remain gravitationally stable, which occurs when $Q_0/12 \gtrsim 1$. For example, for the fiducial values of h, α, and $r_{\rm ODZ}$ in Eq. (3.20), stability is guaranteed as long as $\dot{M} \lesssim 5\times10^{-7}\,M_\odot{\rm yr}^{-1}$. If the outer DZ edge is instead located at 30 AU, $\dot{M} \lesssim 10^{-7}\,M_\odot{\rm yr}^{-1}$ is required for stability.

We note that disk self-gravity may affect the RWI even when $Q > 1$. It has

been shown that the RWI can be suppressed or modified by self-gravity when $Q \lesssim \pi h^{-1}/2$ (e.g., Lovelace & Hohlfeld 2013; Zhu & Baruteau 2015). A self-consistent treatment of self-gravity is required to fully assess its importance relative to RWI in the evolution of DZs.

3.5 Discussion

3.5.1 Summary of Results

We have performed long-term, two dimensional hydrodynamic simulations of protoplanetary disks with dead zones (DZs), modeled as regions with reduced α-viscosities. We give significant attention to the case of narrow DZs, with radial extent of the order or less than the distance to the central star. We found that the vortices and density waves produced by the Rossby wave instability (RWI), triggered at vortensity gradients naturally arising in a DZ, are capable of partially reviving it. The disk eventually reaches a quasi-steady state, in which angular momentum transport and accretion through the DZ are achieved by a combination of the density bump that induces RWI, and Reynolds stress created by waves and vortices. Because of the latter, the Toomre Q parameter, which determines gravitational stability, is reduced with respect to a constant-α disk by at most a factor of 12, i.e., $Q \gtrsim Q_0/12$ [see Eq. (3.20)]. Therefore, RWI can be activated and transport angular momentum through the DZ before it becomes gravitationally unstable, unless the accretion rate is very high. This results in steady accretion, rather than the episodic outburst cycles which may occur when gravito-turbulence transports angular momentum in the DZ (e.g.,

Zhu et al. 2010a, 2010b; Martin & Lubow 2011, 2014). Our results presented in this paper suggest that such episodic cycles are possible only for high accretion rates, e.g., $\dot{M} \gtrsim 5 \times 10^{-7}\, M_{\odot}\mathrm{yr}^{-1}$, for the fiducial disk parameters adopted in Eq. (3.20).

We systematically explored the parameters describing the geometry of the DZ and the disk properties, and quantified the transport efficiency of the RWI and the mass enhancement in the DZ (see Table 3.1). In narrow DZs [$\Delta_{\mathrm{DZ}} \lesssim 0.7$, see Eq. (3.18)], provided that the width of the viscosity transition is not much wider than the local scale height, the azimuthally-averaged Reynolds stresses reach maximum values of $\sim 0.01 - 0.06$, and DZ-averaged Reynolds stresses are in the range $\sim 0.01 - 0.05$. These effective α parameters approximately scale with the active zone viscosity parameter α_0, with the largest values corresponding to the largest viscosity, $\alpha_0 = 0.1$. Typically, the mass in the DZ is enhanced relative to a constant-α disk by a factor of two or less, while the density enhancements in the vortices can be as large as large as 4. For wide DZs, RWI is less efficient, resulting in peak Reynolds stresses $\lesssim 0.02$, and DZ-averaged Reynolds stresses in the range $0.002 - 0.009$. In this case, the mass enhancements in the DZ are about $5 - 7$, with maximum density enhancements reaching $10 - 14$ (relative to constant-α disks).

The morphology of the disk in the quasi-steady state depends on the size of the DZ. For wide DZs, an RWI-unstable vortensity profile is created at either DZ edge, and the resulting vortices merge to produce an $m = 1$ morphology. More interesting phenomena arise for narrow DZs. When the inner and outer DZ edges are sufficiently close, the entire DZ behaves as a single instability site, producing coherent global oscillations. For low viscosities (Reynolds number $\gtrsim$

1800 – 2000), the global wave pattern reaches a nonlinear $m = 1$ mode, consistent with the nonlinear outcome of RWI in the inviscid limit. Increasing the viscosity results in a nonlinear $m = 2$ mode, with two vortices situated at the same radius, separated by 180° in azimuth.

3.5.2 Limitations and Prospects

In our simulations, we modeled the DZ using a radially-dependent viscosity parameter $\alpha(r)$ in a two-dimensional (height-integrated) disk. However, real protoplanetary disks and DZs have three-dimensional, vertically-layered structure (Gammie 1996; Armitage 2011). In a layered disk, MRI-active surface layers of approximately constant surface density $\Sigma_{\rm active} \sim 100\,{\rm g/cm^2}$ bookend a dead midplane with surface density $\Sigma_{\rm tot} - \Sigma_{\rm active}$ that varies with radius. The outer edge of the DZ correponds to the radius beyond which $\Sigma_{\rm tot} < \Sigma_{\rm active}$, so that the entire column of the disk is MRI-active. Thus, a more realistic description of the DZ involves a viscosity parameter which varies with both radius and height, $\alpha(r, z)$. Since the RWI leads to large fluctuations in surface density, a further refinement would require a parameterization of the form $\alpha(r, z, \Sigma)$, in order to account for fluctuations in the depth of the dead layer. These provide possibilities for future three-dimensional simulations.

We have neglected the role of dust grains, which have several important effects. The accumulation of marginally coupled dust grains (those with drag stopping times similar to the dynamical time) in anticyclonic vortices can aid the formation of planetesimals. Vortices in the outer $\sim 50-100$ AU of protoplanetary disks may be responsible for the asymmetries seen in mm-wave observations of

transition disks (e.g., van der Marel et al. 2013; Casassus et al. 2013; Isella et al. 2013), since marginally coupled grains at these radii correspond to a size of mm-cm, and thus contribute significantly to mm emission. Vortices may arise due to the presence of DZs, as described in this paper, if there are viscosity transitions at these large radii. Alternatively, they may be produced by gaps opened by massive planets. If viscosity transitions are present in the inner $\sim$ 10 AU, structures similar to the ones found in our simulations may be found with future observations which can resolve these scales. The gas morphologies presented in this paper cannot be used as proxies for the dust morphologies, but provide a rough approximation. Simulations of mutually coupled gas and dust (e.g., Méheut et al. 2012c; Zhu & Stone 2014) applied to the DZ scenario are required to fully assess this problem.

CHAPTER 4

LONG-LIVED DUST ASYMMETRIES AT DEAD ZONE EDGES IN PROTOPLANETARY DISKS

4.1 Introduction

A number of transitions disks (protoplanetary disks with central dust cavities; see Espaillat et al. 2014 for a review) exhibit significant asymmetries in mm/sub-mm dust emission (van der Marel et al. 2013; Casassus et al. 2013; Isella et al. 2013; Pérez et al. 2014; van der Marel et al. 2016). It is commonly suggested that these asymmetries are the result of dust trapping in large-scale vortices (e.g., Regály et al. 2012; Lyra & Lin 2013; Zhu & Stone 2014). These can arise at axisymmetric "bumps" in the disk, as a result of the Rossby Wave Instability (RWI; Lovelace et al. 1999; Li et al. 2000, 2001; Méheut et al. 2012a). Note that an alternative origin for asymmetries, resulting from the presence of a central binary, was recently suggested by Ragusa et al. (2017).

One route to vortex formation is the opening of a gap by a planet embedded in the disk. The outer edge (and sometimes inner edge) of the gap can be RWI-unstable, resulting in the formation of a vortex (Li et al. 2005). The survival of the vortex is inhibited by both viscosity, and by feedback drag exerted on the gas by the accumulated dust (Fu et al. 2014a, 2014b), although continuous accretion can help to sustain them. Unless both the disk viscosity and dust-to-gas ratio are very small, a vortex formed in this fashion is unlikely to survive for more

This chapter is adapted from Miranda et al. (2017).

than a few thousand orbits, after which the vortex is dissipated, and the dust contained within it is dispersed into a ring (Fu et al. 2014b; Surville et al. 2016). For the typical orbital periods of observed disk asymmetries, this vortex lifetime corresponds to a small fraction of the disk lifetime, which may be several times 10^4 orbits. This implies that it is unlikely to be observed, which is problematic if disk asymmetries are common.

There is an alternative channel for producing dust-trapping vortices via the RWI that does not require embedded planets. Protoplanetary disks are expected to be inefficient at transporting angular momentum via turbulence in their inner regions, due to suppression of the magnetorotational instability by non-ideal MHD effects (Gammie 1996; Bai & Stone 2013; Bai 2014; Lesur et al. 2014; Simon et al. 2015; Bai 2016), and a lack of non-thermal ionization sources (e.g., Cleeves et al. 2013). This region, known as the dead zone (DZ), may extend to a significant fraction of 100AU from the central star. Beyond the DZ, the disk is turbulent, resulting in angular momentum transport and mass accretion. The edge of the DZ is therefore characterized by a sharp increase in turbulent viscosity (for the case of an Ohmic DZ, the gradient of effective viscosity at the DZ edge is sharp, even when that of the underlying resistivity is not; Lyra et al. 2015). The presence of a viscosity transition leads to accumulation of mass, creating an RWI-unstable bump at which a vortex may be formed (Varnière & Tagger 2006; Lyra et al. 2009; Regály et al. 2012; Lyra & Mac Low 2012; Miranda et al. 2016). Note that the DZ also has an inner edge, but its proximity to the central star ($\sim$ 0.1AU; Gammie 1996) makes it irrelevant to the dynamics of the outer disk ($\sim$ 50AU) where asymmetric features are observed, and so we are only concerned with the outer edge of the DZ in this study.

In this paper, we carry out high-resolution 2D hydrodynamic simulations of vortex formation at a DZ edge, including a full treatment of dust dynamics, with feedback. We evolve the disk for 10^4 orbits, representing a substantial fraction of the lifetime of a protoplanetary disk. We find that, unlike in the case of a vortex at a planetary gap edge, non-axisymmetric dust trapping can be maintained for very long periods of time. We also carry out radiative transfer calculations to produce simulated sub-mm images from our results.

The outline of this paper is as follows. In Section 4.2, we describe the setup for our high-resolution numerical simulations, as well as the method for creating simulated mm images from them. In Section 4.3, we present the main results of the numerical simulations, as well as the simulated images. Finally, in Section 4.4, we discuss and contextualize our results.

4.2 Numerical Setup

We consider a two-dimensional thin disk of gas and dust described in polar coordinates (r, ϕ) by gas surface density Σ_g, dust surface density Σ_d, gas velocity $\mathbf{v}_g$ and dust velocity $\mathbf{v}_d$, around a star of mass M_*. The equation of state for the gas is locally isothermal, $P = c_s^2(r)\Sigma_g$, where P is the height-integrated pressure and $c_s(r) = c_{s,0}(r/r_0)^{-1/4}$ is the radially-dependent sound speed. The scale height of the disk is $H = c_s/\Omega_K$, where $\Omega_K = (GM_*/r^3)^{1/2}$ is the Keplerian orbital frequency. The value of $c_{s,0}$ is chosen so that $H/r = 0.05$ at r_0, and the disk is slightly flared, with $H/r \propto r^{1/4}$.

We solve the two-fluid hydrodynamic equations describing the coupled evolution of gas and dust using the LA-COMPASS code (Li et al. 2005, 2009; Fu et

al. 2014a, 2014b), including the effects of aerodynamic drag on the dust, as well as on the gas (i.e., the back-reaction or feedback), and dust diffusion. The radial extent of the computational domain is $[0.2, 4.39]$, in units of the scaling radius r_0. The initial surface density profie is $\Sigma_g = \Sigma_0(r/r_0)^{-1}$, which corresponds to a steady state for a disk with constant α and our chosen sound speed profile. Initially, the dust surface density follows the gas surface density according to $\Sigma_d = \eta_d \Sigma_g$. Our standard value for the initial dust-to-gas ratio is $\eta_d = 0.01$. The value of this parameter, along with several others described in this section, were varied in several parameter study runs, which are defined in Table 4.1. The gas surface density and velocity are fixed at their initial values at both the inner and outer boundaries. An outflow inner boundary condition and a zero radial velocity outer boundary condition are imposed on the dust. We do not consider the effects of disk self-gravity or the "indirect potential" due to the motion of the central star (this is justified by our small disk mass, $M_{\rm disk}/M_* \approx 10^{-4}$; see Section 4.2.2). Throughout this paper we express time in units of "orbits", where 1 orbit $= 2\pi/\Omega_0$ is the Keplerian orbital period at r_0.

Our numerical resolution is $N_r \times N_\phi = 2048 \times 3072$ (with uniform grid spacing in both r and ϕ), so that the disk scale height is resolved by about 25 grid cells at r_0, and each run is evolved for 10^4 orbits. The Standard run was also simulated at a higher resolution, $N_r \times N_\phi = 4096 \times 6144$, for 4000 orbits, for comparison.

4.2.1 Viscosity Profile

The gas has a (turbulent) kinematic viscosity given by $\nu = \alpha c_s^2/\Omega_K$ (Shakura & Sunyaev 1973). The DZ is modeled using a radially varying viscosity parameter,

$$\alpha(r) = \alpha_0 \left\{ 1 - \frac{1}{2}\left(1 - \frac{\alpha_{DZ}}{\alpha_0}\right)\left[1 - \tanh\left(\frac{r - r_{DZ}}{\Delta_{DZ}}\right)\right]\right\}. \tag{4.1}$$

Here $\alpha_0 = 10^{-3}$ and $\alpha_{DZ} = 10^{-5}$ are the active zone and dead zone viscosity parameters (these values are broadly consistent with recent non-ideal MHD simulations, e.g., Simon et al. 2015, and models motivated by such simulations, e.g., Bai 2016). Note that our results are not strongly dependent on the value of α_{DZ}, as long as it is much less than α_0. The transition between the two regions is located at $r_{DZ} = 1.5r_0$, and has a width of Δ_{DZ}, which in our standard run is chosen to be equal to the local scale height, $\Delta_{DZ} = H(r_{DZ}) = 0.083r_0$.

The viscous timescale is (e.g., Lynden-Bell & Pringle 1974) $t_{visc} = (4/9)r^2/\nu = 2.8 \times 10^4 (r/r_0)$ orbits (for $r/r_0 \gtrsim 1.5$). Therefore, the 10^4 orbit duration of our simulations is not long enough for significant global evolution of the gas surface density profile to occur. The viscous timescale associated with the viscosity transition, i.e., the timescale for bump accumulation, is smaller by a factor of $(\Delta_{DZ}/r)^2$, resulting in $t_{bump} \approx 100$ orbits.

4.2.2 Gas and Dust Parameters

We adopt the observationally derived parameters for the disk around Oph IRS 48, choosing $M_* = 2M_\odot$, $r_0 = 50$AU (so that 1 orbit = 250 yr), and $\Sigma_0 = 5.401 \times 10^{-6} M_*/r_0^2 = 0.0384 \mathrm{g/cm^2}$ (Bruderer et al. 2014). Our choice of r_0, along with our adopted DZ parameters (edge at 75AU), result in vortices that

form at about 60 – 65AU from the central star, coincident with the location of the peak mm emission in the system. The dust dynamics are characterized by the Stokes number, which, in the Epstein regime, is given by $\mathrm{St} = \pi s_\mathrm{p} \rho_\mathrm{p} / (2\Sigma_\mathrm{g})$, where s_p and ρ_p are the dust size and density. We choose $\rho_\mathrm{p} = 0.8 \mathrm{g/cm^3}$ and $s_\mathrm{p} = 1\mathrm{mm}$, resulting in $\mathrm{St} = 3.27$ at r_0 initially. Thus, the dust experiences nearly the maximum possible radial drift (the maximum occurs for $\mathrm{St} = 1$), and is also nearly the most susceptible to being trapped in vortices, hence giving the strongest possible feedback to the gas.

The total gas mass is $M_\mathrm{g} = 0.30 M_\mathrm{J}$ and the total dust mass is $M_\mathrm{d} = 0.94 M_\oplus$, although these values are sensitive to the extent of the disk, especially the location of the outer boundary, which is somewhat arbitrary. For the canonical parameters, the time required for dust to drift from r_out to $r = 1.25$, the approximate location where vortices form, is $t_\mathrm{drift} = 1250$ orbits. There is about $0.71 M_\oplus$ of dust beyond the vortex radius, which could potentially be trapped there. As the duration of our simulation is about 8 times the drift timescale, we expect that the dust will have fully settled into a global quasi-equilibrium by the end of the simulation, i.e., drift on a global scale will no longer be occurring.

4.2.3 Synthetic Observations

We use the results of our hydrodynamic simulations to produce simulated maps of dust continuum emission, using RADMC-3D (Dullemond 2012), following the method described by Jin et al. (2016), which we briefly summarize here. First, we perform a thermal Monte Carlo simulation to determine the temperature structure of the disk, which is determined by the distribution of small (μm) dust

grains. For our standard parameters, μm-sized dust grains have St $\sim$ 0.003 at r_0, and so do not necessarily remain perfectly coupled to the gas over thousands of orbits. Nonetheless, we assume that their distribution follows the three-dimensional structure of the gas, which is extrapolated from the gas surface density, with a fixed dust-to-gas ratio. The dust opacity for μm-sized dust used in the temperature calculation is modeled as in Isella et al. (2009), using a grain size distribution $n(a) \propto a^{-3.5}$ between 0.005μm and 100μm. The disk is illuminated by a star with $T_{\rm eff}$ = 9500K and $L = 24L_\odot$ (using the parameters for IRS 48 from Follette et al. 2015). Next, the two-dimensional surface density of large (mm) dust is also extrapolated into a three-dimensional density. These particles generally have a different scale height than the gas due to vertical settling. If settling is opposed only by turbulent diffusion, the dust scale height is $H_{\rm d} = \sqrt{\alpha/({\rm St}+\alpha)}H_{\rm g}$ (e.g., Birnstiel et al. 2016), which can be extremely small compared to the gas scale height. However, in the scenario we consider, vertical motions associated with vortices may not allow dust to settle to this degree. Instead, we adopt $H_{\rm d} = 0.1H_{\rm g}$, based on the results for vertical dust distributions in vortices from Méheut et al. (2012b), for the largest particles they considered (St = 0.5). Finally, the density distribution of mm dust, along with its opacity (calculated the same way as for the μm dust, but with a maximum grain size of 1mm), and the temperature distribution, are used to create a synthetic thermal emission map at 440μm (ALMA Band 9). The system is placed at a distance of 120pc with an inclination of 50° (again adopting the parameters of IRS 48), and the emission map is convolved with a Gaussian beam of a given size in order to simulate an interferometric observation.

Name	Σ_0 [g/cm^2]	η_d	Δ_{DZ}
Standard	0.0384	0.01	H
GasHigh	0.3840	0.01	H
DustHigh	0.0384	0.10	H
ViscSharp	0.0384	0.01	$H/2$
ViscBroad	0.0384	0.01	$2H$

Table 4.1: Simulation names and parameters

4.3 Results

4.3.1 Standard Run

Figure 4.1 summarizes the evolution of the Standard run (see Table 4.1). Here, snapshots of gas surface density, dust surface density, and Rossby number (dimensionless vorticity perturbation) $\mathrm{Ro} = [\nabla \times (\mathbf{v}_g - \mathbf{v}_K)]_z/(2\Omega_K)$ are shown at several key points in time. Note that negative values of Ro correspond to local rotation in the opposite direction of the bulk orbital motion, i.e., anticyclonic rotation (only anticyclonic vortices are stable against the Keplerian shear of the disk). In the first snapshot ($t = 300$ orbits), three vortices have been formed by the RWI, which has been triggered at the bump formed by accumulation of mass near the viscosity transition. A significant amount of dust has been collected axisymmetrically at the bump, such that $\Sigma_d/\Sigma_g \approx 1$. In the next snapshot (500 orbits), the three gas vortices have merged into a single vortex. The ring of dust at the pressure maximum has become clumpy, with $\Sigma_d/\Sigma_g \approx 10$ at some points. In the third snapshot (1000 orbits), the configuration is not much different than the previous one (although by this point, the outer disk has been entirely cleared of dust due to radial drift). In the fourth snapshot (2000 orbits), the amplitude of the gas density perturbation has increased due to continued

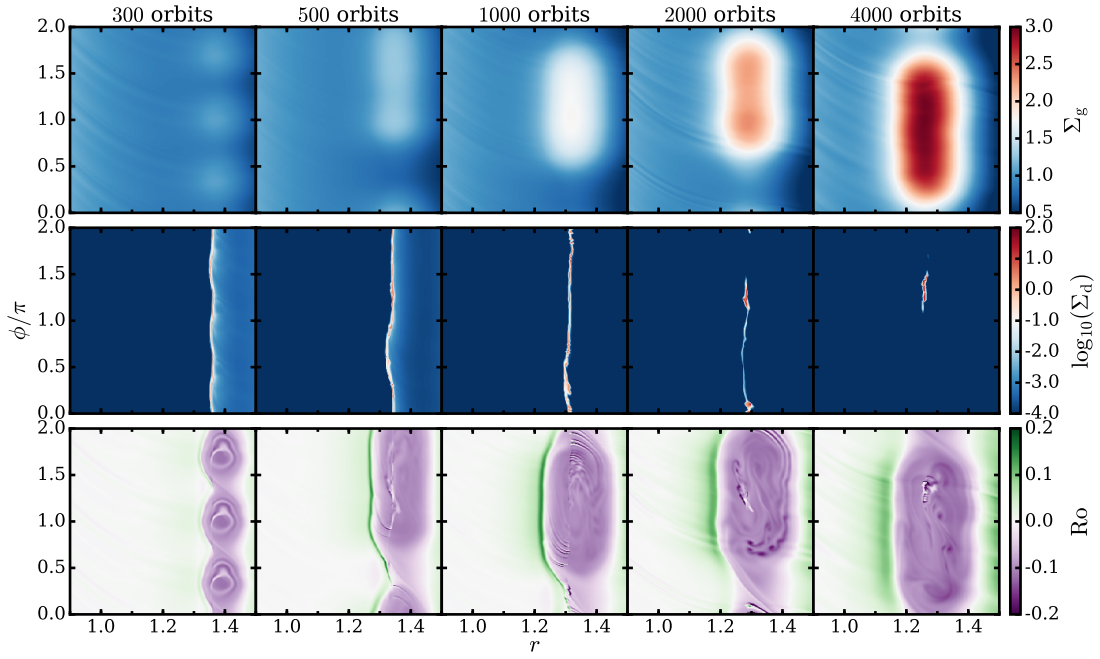

Figure 4.1: Snapshots of gas surface density Σ_g (in units of Σ_0; top row), dust surface density Σ_d (also in units of Σ_0; middle row), and dimensionless vorticity or Rossby number, Ro (bottom row), in the $r - \phi$ plane, for the Standard run. Here the full azimuthal extent of the disk is shown, but only a small fraction of the radial domain, centered on the vortex region, is shown. Note that Σ_d is shown on a logarithmic scale.

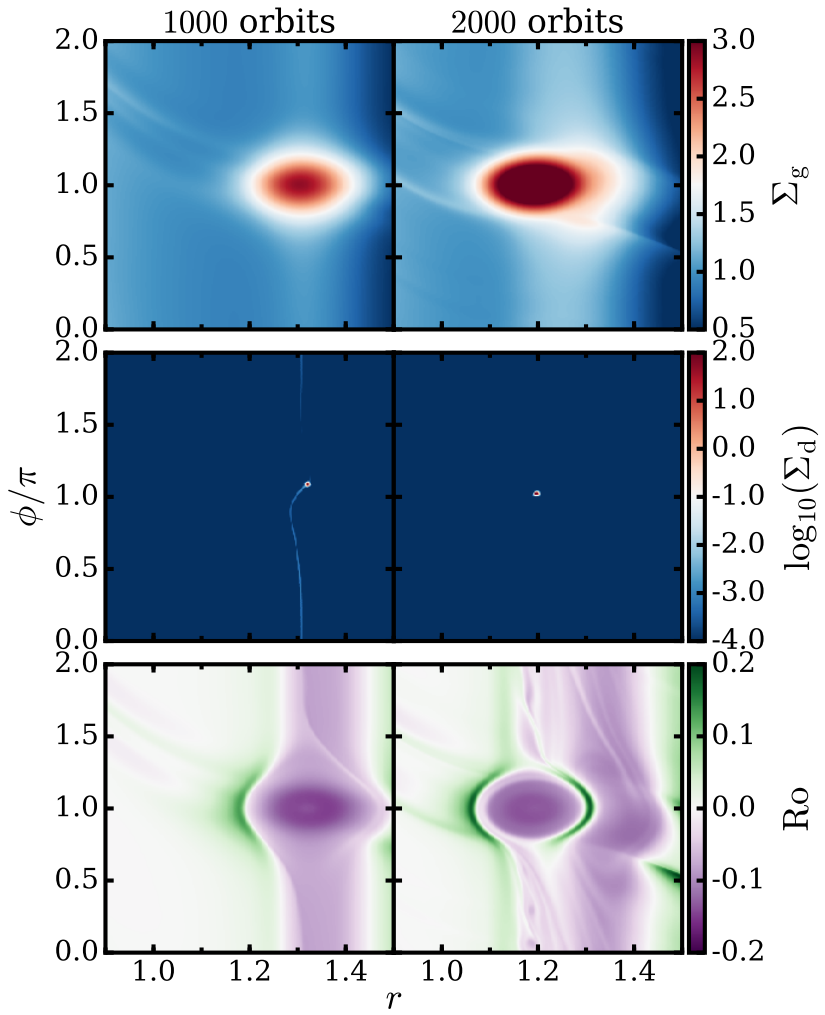

Figure 4.2: Same as Figure 4.1, with dust feedback turned off.

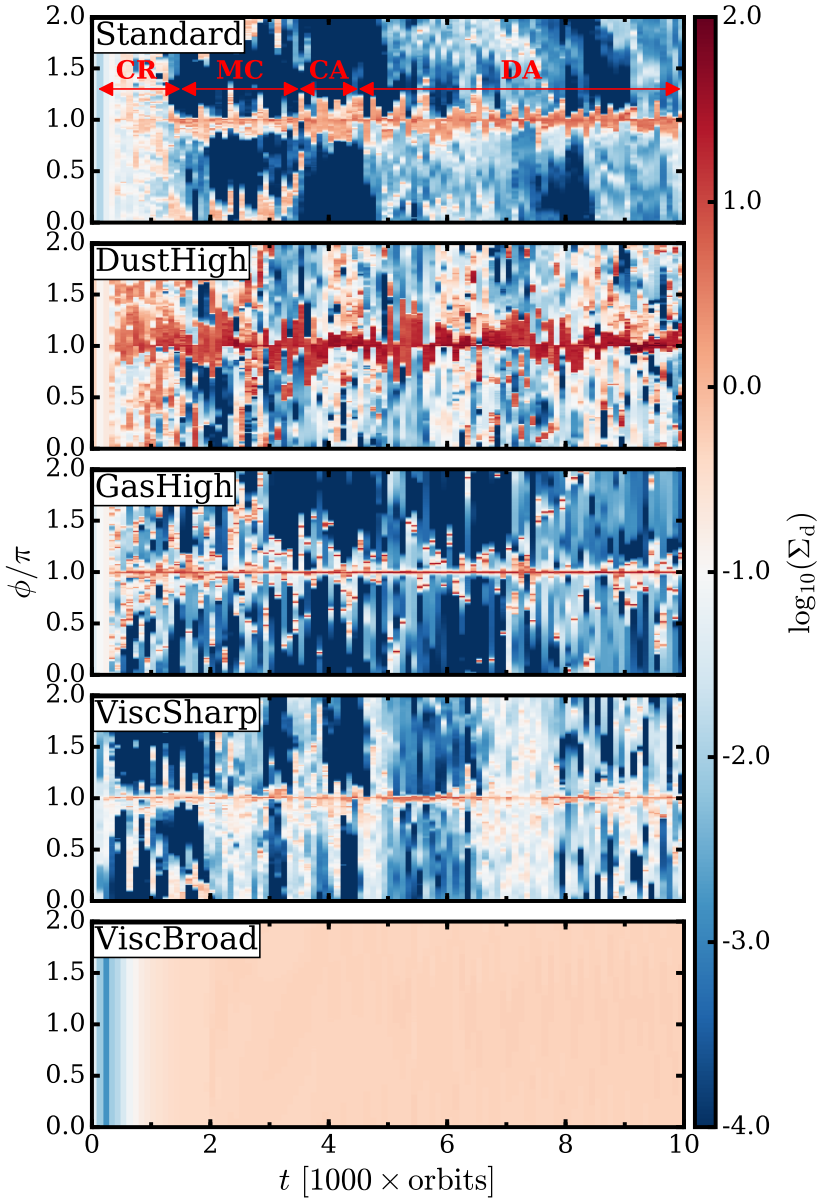

Figure 4.3: Azimuthal dust distribution as a function of time, for different runs. In each panel, vertical slices represent the average dust density distributions in an annulus of width $H/2$, centered on the grid cell containing the largest dust density, at different times (the profiles are shifted so that the maximum coincides with $\phi = \pi$). For the Standard run, the time intervals corresponding to the clumpy ring ("CR"), multiple clumps ("MC"), clean asymmetry ("CA"), and dirty asymmetry ("DA") phases are indicated.

accumulation of gas from the viscosity transition. At this point, much of the dust in the ring has become concentrated into two clumps, which are separated in azimuth. Finally, in the fifth snapshot (4000 orbits), the dust has become mostly concentrated into a single compact clump, in which the dust-to-gas ratio is in excess of 100 (although the location of the largest dust density is not exactly coincident with that of the largest gas density, which is also true in some of the previous snapshots). Additionally, the peak gas density has modestly increased since the last snapshot ($\sim$ 20% in 2000 orbits), due to continued accretion. Note that at later times, the dust becomes somewhat more spread out again, although its distribution still remains strongly non-axisymmetric.

Figure 4.2 shows several snapshots of the Standard run with feedback turned off, in order highlight its effects. Comparing with Figure 4.1, we see that feedback weakens the gas vortex, i.e., reduces the amplitude of the gas density and vorticity perturbations, and causes it to become elongated. Without feedback, the vortex aspect ratio (determined by the shape of the Ro = 0 contour bounding the region of anticyclonic vorticity) is $\sim$ 8 at 2000 orbits, with feedback it is $\sim$ 20. Since the vortex is much stronger without feedback, it migrates substantially (Li et al. 2001; Paardekooper et al. 2010), moving inward by about 5AU in 1000 orbits. Thus, feedback weakens the vortex so that it migrates negligibly, staying in nearly the same place for 10^4 orbits. The "turbulent" features in the vortex, traced by the vorticity field, are caused by the feedback, as without it, the vorticity distribution is much smoother. The "shock" feature at which the vorticity changes sign, effectively defining the edge of the vortex, is present regardless of whether or not feedback is turned on. Without feedback, the dust is very efficiently accumulated in the vortex, collecting into what is essentially a single point by 1000 orbits, and remaining that way indefinitely. This is in

stark contrast to the case with feedback, for which the dust is still in the form of a clumpy ring at 1000 orbits, and does not accumulate into a single feature (which is more extended than in the no feedback case) until 4000 orbits. Feedback therefore slows down and reduces the efficiency of dust trapping, but does not completely inhibit it.

The evolution of the azimuthal dust distribution in the vortex region for the Standard run (with feedback) is further illustrated in Figure 4.3 (top panel). We use this to define several different phases or morphologies that the dust can exhibit. Between a few hundred and about 2000 orbits, the dust distribution is not completely axisymmetric, but rather is clumpy. We denote this the "clumpy ring" phase, which is represented by the first three columns of Figure 4.1. Between about 2000 and 4000 orbits, in what we denote the "multiple clumps" phase, which illustrated by the fourth column of Figure 4.1, there are two distinct clumps, initially separated by $\sim 180^{\circ}$. The two clumps eventually merge, leading to the "clean asymmetry" phase (e.g., the fifth column of Figure 4.1), which persists for about 1000 orbits. In this phase, essentially all of the dust in the annulus is contained in this clump. For the remainder of the evolution (after about 5000 orbits), there is still one strong feature, but it is partially surrounded by a residual ring with a lower surface density. We denote this the "dirty asymmetry" phase. A snapshot of this phase, which appears to be the typical state for many thousands of orbits, is shown in Figure 4.4, at 10^4 orbits. Note that the distinction between the different morphologies is somewhat subjective and not concrete. For example, two discrete clumps of dust closely separated in azimuth tend to appear as a single feature in interferometric images (see Section 4.3.3), and so there exists a continuum between the multiple clumps and a clean asymmetry morphologies. Nonetheless, this classification scheme is useful for

characterizing our results.

In addition to the asymmetric features we have focused on, there is also an axisymmetric ring of dust at about $r = 2.4$ (120AU). The ring forms as a result of a weak pressure maximum that forms there, in response to the pressure minimum formed just outside of the DZ edge. Once the ring is formed, dust that drifts from the outer disk gets trapped in the ring and does not make it to the vortex region. Thus, there is a finite supply of dust that can be trapped in the vortex region. A larger computational domain with more dust available in the outer disk would result in more dust being trapped in the ring, but the vortex region would be unaffected. Together, the ring beyond the DZ edge and the asymmetric features interior to it will make up the main features of our simulated images (see Section 4.3.3).

We investigate the numerical convergence of our results in Figure 4.5, which shows the maximum dust-to-gas ratio during the initial evolution of the Standard run, for several different resolutions. In all cases there is an approximately exponential rise as dust is rapidly collected in the bump/vortices, followed by fluctuations around an approximately constant level, which is always in excess of unity. Larger maximum values are achieved with increasingly higher resolution. The range of the fluctuations is similar for the two highest resolutions (2048×3072 and 4096×6144), while for the lowest resolution (1024×1536), the dust-to-gas ratio typically remains significantly below this range. However, even for the highest resolution, for which case the dust-to-gas ratio can reach several hundred, and thus feedback is very strong, the overall results are unchanged: the dust exhibits significant azimuthal asymmetries for the duration of the simulation.

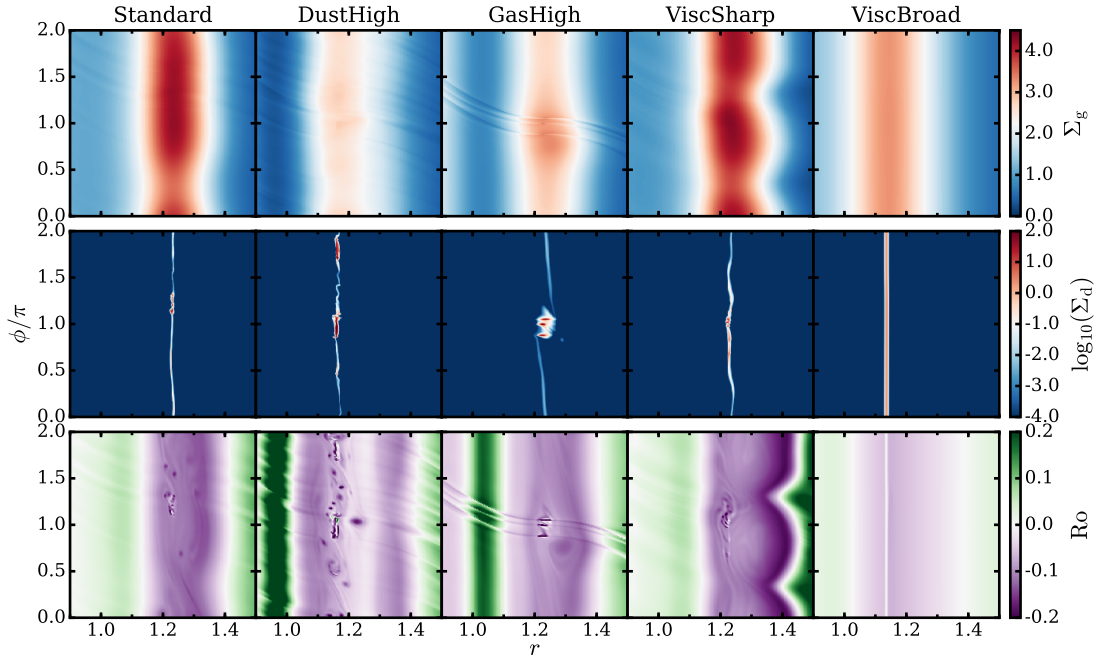

Figure 4.4: Snapshots of different runs at 10^4 orbits.

4.3.2 Parameter Dependence

We investigated the effects changing several of the parameters of the Standard run. The parameters for these runs are given in Table 4.1. Feedback is included in all of these runs. The main results are shown in Figure 4.3, which illustrates the evolution of the azimuthal dust distribution in the vortex region, and Figure 4.4, which shows snapshots at 10^4 orbits.

Dust-to-Gas Ratio

A number of transition disks have been found to have enhanced dust-to-gas ratios, as large as 0.1, compared the primordial disks, for which the typical value is 0.01. We performed a simulation with $\eta_d = 0.1$ (10 times larger than in the Standard run), labeled "DustHigh". Note that enhancement occurs in part due

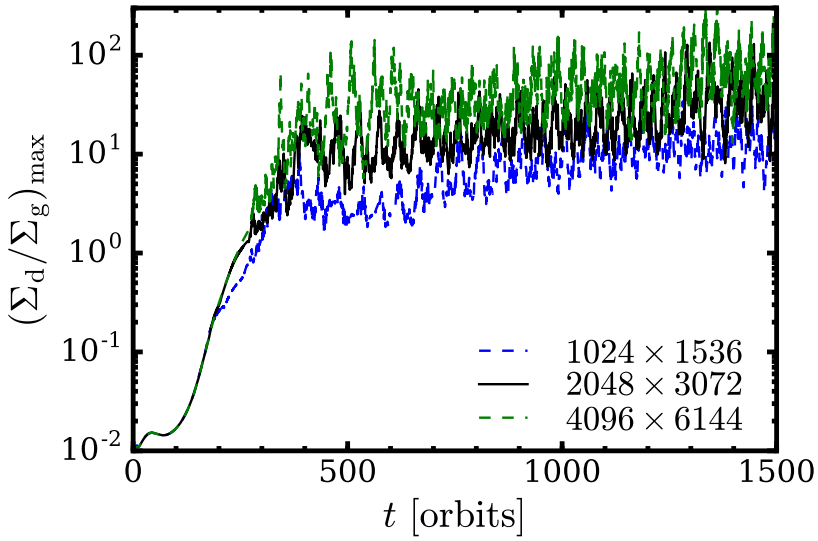

Figure 4.5: The maximum dust-to-gas ratio versus time for the Standard run, for several numerical resolutions.

to radial drift of dust, while the initial condition in our simulation essentially represents a disk for which no drift has occured. Thus, there is not necessarily a direct correspondence between an observed dust-to-gas ratio and our value of η_d.

As there is a larger reservoir of dust mass available to be trapped by the vortex, larger maximum dust surface densities are reached (a few times larger than in the Standard run). As a result, dust feedback is stronger, making dust concentration more difficult. This is evident in the azimuthal dust profile shown in Figure 4.3: the "clean asymmetry" is only seen briefly, around 3500 – 4000 and 5000 – 6000 orbits. Instead, the distribution can usually be described by the multiple clumps or dirty asymmetry morphology. Nonetheless, the dust distribution remains asymmetric at all times.

Disk Mass

The run labeled "GasHigh" features a gas surface density 10 times larger than in the canonical run (the dust surface density is also larger so that the standard value of $\eta_{\rm d}$ is maintained). As the standard gas surface density profiles represents IRS 48 as it is observed today, at an age of 8Myr, after experiencing significant evolution (e.g., due to spreading, accretion, disk winds, photoevaporation, etc.), this profile could represent the system at an earlier time when the disk was more massive. The initial Stokes number of 1mm dust at r_0 is 0.327 in this run, and so the dust dynamics are different than in the Standard run. Note that this run can also represent the dynamics of particles ten times smaller (0.1mm) with the standard gas density profile.

Owing to the different dust dynamics, the dust is trapped in the vortex differently, typically forming several (2 – 4) compact clumps, usually separated by not more than 180° in azimuth, and often much less, as seen in Figure 4.3. The clean or dirty asymmetry morphologies are rarely seen. The most striking difference compared to the Standard run is that the axisymmetric dust ring at $2.4r_0$ is not formed. Since the dust in the outer disk has St $\approx$ 1, the time required for dust to drift from the outer boundary to $r \approx 1.25r_0$ is only about 300 orbits, about 4 times faster than in the Standard run. This is shorter than the time required to form the weak secondary pressure maximum beyond the DZ edge, thus dust never has a chance to be trapped there, and can only be trapped by the vortex. This leads to a situation in which nearly all of the dust in the disk is situated on one side of the star at at any given time.

Viscosity Transition Width

We vary the width of the viscosity transition, in the runs labeled "ViscSharp" and "ViscBroad", which have $\Delta_{DZ} = H/2$ and $2H$, respectively. In the ViscSharp run, the sharper viscosity gradient leads to more vigorous accumulation of mass at the DZ edge, which replenishes the RWI-unstable bump more quickly, resulting in more robust vortex formation. In the snapshot shown in Figure 4.4, there are actually two vortices present, indicating that they are continuously being formed (i.e., RWI modes with $m > 1$ are still being excited). The vortex is therefore less susceptible to destruction by feedback. The amplitude of the gas perturbation is larger (although it is still not strong enough to migrate substantially), and most of the dust tends to be maintained in the center of a vortex. The evolution of the azimuthal dust distribution is not very different from the Standard run, except that the dirty asymmetry morphology is dominant most of the time.

In the ViscBroad run, mass is not accumulated into a sharp enough bump to trigger the RWI, so the disk remains axisymmetric (this result is in agreement with previous studies which find that excitation of the RWI requires $\Delta_{DZ} \lesssim 2H$, e.g., Lyra et al. 2009; Regály et al. 2012). However, at around 1800 orbits, there is a transient growth of a very small deviation from axisymmetry in the gas that lasts for 200 orbits and reaches an amplitude of only $\sim 1\%$, which is then damped. In response, the dust becomes slightly asymmetric (barely discernible in Figure 4.3), although the asymmetry decays with time. Visually, the dust distribution near the DZ has a ring morphology for the entire simulation. Additionally, in this run, while a ring instead of an asymmetry is formed at $r \approx 1.25$, the second ring at $r \approx 2.4$, which appears in the other runs, is not created. The

reason for this is similar to the reason for the lack of a ring in the GasHigh run– dust drifts through the outer disk before the secondary bump is created, in this case due to the increased timescale for gas surface density evolution, owing to the broader viscosity transition.

Initial Surface Density Profile

In all of the runs we have presented, we adopted smooth initial profiles for the gas and dust surface densities. An RWI-unstable bump is self-consistently produced by the DZ viscosity transition, as the disk attempts to reach an equilibrium in which the accretion rate, $\dot{M} \sim \nu\Sigma_g$, is radially constant (this requires Σ_g to be enhanced by a factor of $\alpha_0/\alpha_{DZ} = 100$ in the DZ). The smooth initial profile, which ignores the presence of the viscosity transition, may not necessarily represent the true initial state of the disk, and so we may ask how our results are affected if the disk has already had a chance to partially evolve toward this equilibrium. We explored this by first allowing the disk to evolve in 1D (i.e., with enforced axisymmetry) for several thousand orbits, before proceeding to evolve it in full 2D. We find that, as the gas surface density becomes enhanced in the DZ, a sharp bump near the viscosity transition always arises, since the viscous timescale associated with the viscosity transition is much shorter than the timescale to distribute mass farther inwards towards the DZ. Therefore, regardless of the specifics of the global profile, the viscosity transition guarantess the existence of an RWI-unstable bump which forms vortices. Radial drift ensures that dust becomes localized to the bump.

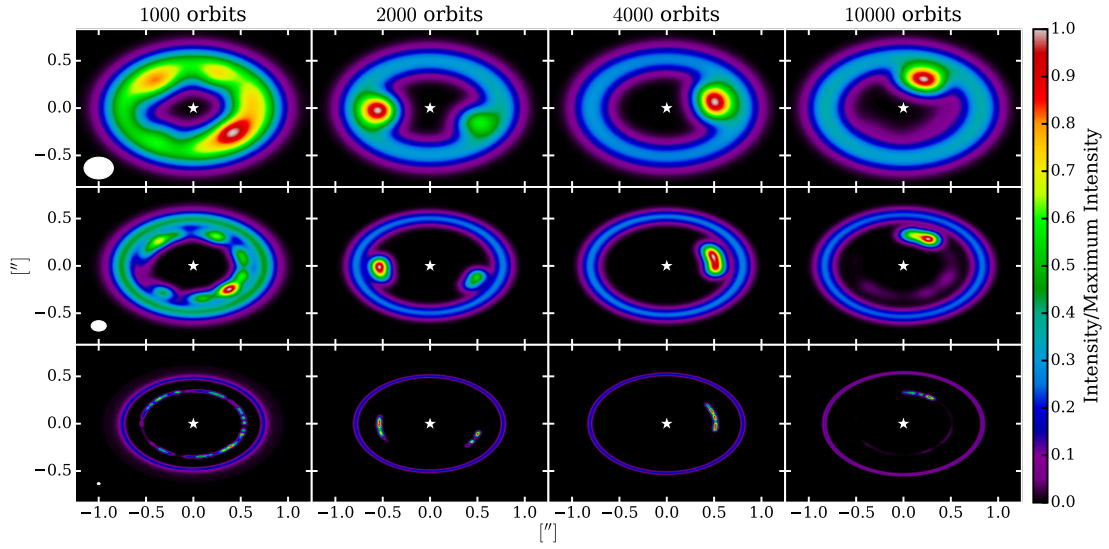

Figure 4.6: Synthetic images at 440μm for the Standard run at several points in time. The disk is placed at a distance of 120pc with an inclination of 50°, and convolved with three different Gaussian beams: 0.31″ × 0.23″ (top row), 0.16″×0.11″ (middle row), and 0.03″×0.02″ (bottom row). The beams are shown in the leftmost panels of each row. The different snapshots correspond to several different possible morphologies, including (from left to right) the clumpy ring, multiple clumps, clean asymmetry, and dirty asymmetry phases. Note that each image has each been scaled by its maximum intensity, which ranges from about 2 – 4 mJy/beam (for the smallest beam size) to about 15 – 20 mJy/beam (for the largest beam size).

4.3.3 Synthetic Images

Images of the different phases of the Standard run are shown in Figure 4.6, for different beam sizes. An ubiquitous feature in all of them is a ring at about 120AU, corresponding to dust trapped at the secondary pressure maximum that arises outside of the DZ edge. The appearance of this ring does not change much during the disk evolution. Additionally, there is the asymmetric feature at about 60−65AU, corresponding to the region shown in Figures 4.1 and 4.3. For this run, the appearance of the dirty asymmetry morphology is not very different from that of the clean asymmetry morphology, except for an arc of emission whose

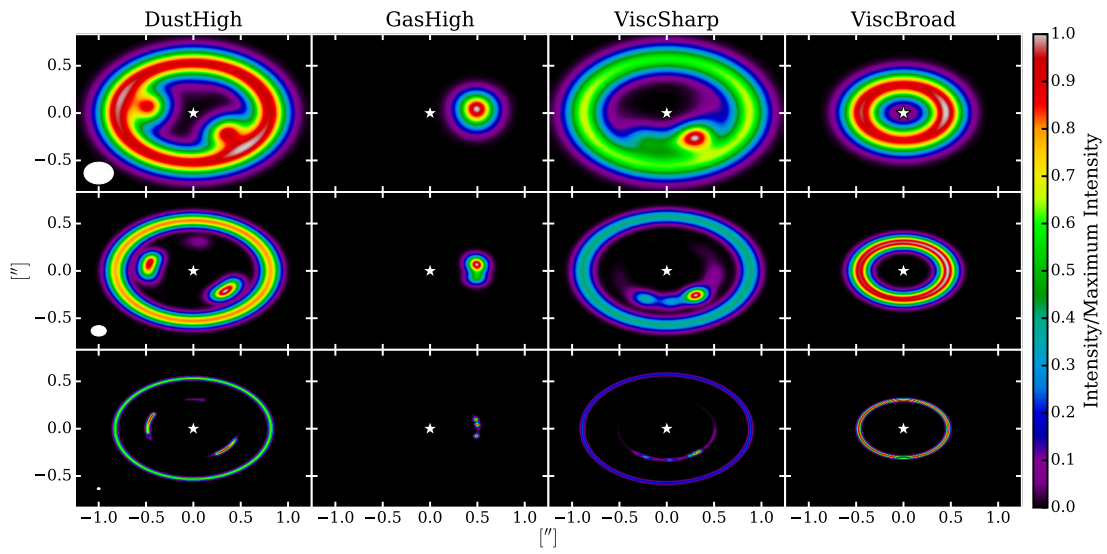

Figure 4.7: Synthetic images at 440μm (as in Figure 4.6) at 10^4 orbits for the runs with varied parameters.

brightness is about 10% of the peak brightness of the asymmetric feature.

In all 4 images, the ring and asymmetric feature are not resolvable as separate features when convolved with the largest beam, which corresponds approximately to the resolution of ALMA cycle 0 (as in the image of IRS 48 in van der Marel et al. 2013). As a result, the ring and asymmetry appear nearly coincident radially, and the brightness contrast around the apparent asymmetric ring appears to be small (a few). For a beam that is twice as small, the ring and asymmetry can be distinguished from one another, and the brightness contrast along the annulus containing the asymmetric feature is revealed to be very large–effectively infinite on the linear scale shown (in fact it is about 10^3). At the highest resolution (corresponding approximately to the optimum resolution achievable by ALMA), it is revealed that some of the asymmetric features, which appear to constitute a single feature at lower resolutions, are in fact made of several compact clumps or arcs. For example, in the image taken at 4000 or-

bits, what appeared to be a single feature at lower resolution is shown to in fact be 4 very small patches of emission clustered together azimuthally. However, as these patches are very small (only a few grid cells wide in the hydro simulations), they are still not resolved at this resolution, and their apparent size is consistent with the beam size.

Figure 4.7 shows images of the other runs taken at 10^4 orbits. The DustHigh run appears similar to the Standard run at 2000 orbits, although the ring is almost as bright as the asymmetric features, as a result of more dust being trapped there. In the GasHigh run, since the outer ring is not formed, the dust emission is dominated by the asymmetric feature, so that all of the dust appears to be on one side of the disk. For this reason, the low resolution image of this run bears the most resemblance to IRS 48. In this case, as for the Standard run, at optimum resolution, the asymmetric feature is revealed to actually consist of several unresolved sub-components. The ViscSharp image is similar to the clean asymmetry (or perhaps the multiple clumps) phase of the Standard run. The ViscBroad run exhibits only an axisymmetric ring (with 10% brightness variations) at the same radius that the other runs have asymmetries. Except for the ViscBroad run, none of the runs are completely axisymmetric at 10^4 orbits.

4.4 Discussion

We performed high-resolution, two-dimensional hydrodynamic simulations of dust trapping in vortices formed at the outer edges of DZs, including the effect of dust feedback. We found that, while feedback somewhat inhibits and slows down the process of azimuthal dust trapping, it does not ultimately prevent

it from occuring. This is in contrast to the case of a vortex at the edge of a planetary gap, in which the vortex is destroyed and the dust is released into a ring. The key difference between these scenarios is that in the DZ case, the viscosity transition leads to continuous accumulation of gas at the DZ edge, which allows the RWI to be sustained, constantly replenishing the vortex. This process is able to overcome dust feedback, which attemps to destroy the vortex. As a result, dust remains trapped, although the vortex is weak, and the trapping is weaker than it is without feedback. Asymmetric features in emission maps of the disk persist for at least 10^4 orbits (the total duration of our simulations). Therefore, a disk with a DZ may appear asymmetric for most of its lifetime. We created synthetic images of thermal dust emission to compare the appearance of these features to those observed in transition disks.

Observed asymmetric features in transition disks appear to have large radial widths, much wider than the coherence width of a vortex, which is no more than a few pressure scale heights. However, the apparent widths are approximately consistent with one beam width, suggesting that the features are not resolved. Future observations may have the potential to resolve these features, which may be much narrower. This has also been proposed for vortices at planetary gap edges (Zhu & Stone 2014), though larger features are possible due to the effects of self-gravity and the reflex motion of the central star, provided the disk is sufficiently massive (Mittal & Chiang 2015; Zhu & Baruteau 2016, Baruteau & Zhu 2016). Additionally, the observed asymmetric features have different azimuthal widths, ranging from relatively compact ($\sim 45°$) to extended "horseshoe/banana" features extending more than $180°$ in azimuth. As with the radial width, the azimuthal extent may also be unresolved. Our results suggest that, if transition disk asymmetries are a result of viscosity transitions

at DZ edges, they consist of very compact clumps of dust, or perhaps multiple clustered clumps, which individually are unresolvable even with a resolution of 0.02''.

Another distinctive feature of our results is that there is also an axisymmetric ring of dust emission, located about twice as far from the star as the asymmetric features. When observed with low resolution, the two features blur together, taking the appearance of an asymmetric ring with only small brightness variations in azimuth. Higher resolution observations allow the two features to be resolved, revealing that the asymmetry in fact has a much higher brightness contrast. A similar phenomenon, in which an apparent weak asymmetry is resolved into a stronger asymmetry and an axisymmetric ring, located at different distances from the central star, was seen in observations of HD 135344B/SAO 206462. The asymmetry was first seen to be quite weak, with an azimuthal intensity variation of less than 2 (Pérez et al. 2014), due to the (relatively) low resolution of the observation. Later observations, which achieved a better spatial resolution of 0.16'', were able to distinguish an asymmetry with an azimuthal intensity variation of 4 from a ring which is very close to axisymmetric, with an intensity variation of less than 1.2 (van der Marel et al. 2016). However, the asymmetric feature is farther from the the star than the axisymmetric ring, in contrast to our results, in which the asymmetric feature is closer to the star than the ring. Additionally, this object also exhibits spiral structure in infrared scattered light, which may indicate the presence of a planet that may be responsible for the asymmetry. Nonetheless, this demonstrates the potential for future observations to distinguish a weak asymmetry from an unresolved ring/asymmetry combination.

A large amount of dust is trapped in the asymmetric features near the DZ edge. A few tenths of an Earth mass are accumulated in the the most dense clumps. As the dust densities reach about $10 g/cm^2$ (or several times larger if the initial dust-to-gas ratio is enhanced), gravitational collapse or the streaming instability may occur in these clumps (see Raettig et al. 2015). They contain enough mass to form not just planetesimals, but potentially planetary embryos directly. Additionally, if such embryos are created, the trapped particles, which experience maximal drift with respect to the gas, may be quickly accumulated further through pebble accretion (e.g., Owen & Kollmeier 2016).

CHAPTER 5

TIDAL TRUNCATION OF INCLINED CIRCUMSTELLAR AND CIRCUMBINARY DISKS IN YOUNG STELLAR BINARIES

5.1 Introduction

Protoplanetary disks are often found in binary systems. In general, because of the complex (and often turbulent) formation processes of stars, binaries, and disks (e.g., Bate, Bonnell & Bromm 2003; McKee & Ostriker 2007; Klessen 2011; Fielding et al. 2015), both circumstellar and circumbinary disks are likely formed with inclined orientations with respect to their binary orbital planes. Indeed, a number of binary young stellar objects (YSOs) are observed to contain circumstellar disks that are misaligned with the binary plane (e.g., Stapelfeldt et al. 1998, 2003; Neuhäuser et al. 2009). The orientation of jets in several unresolved YSOs or pre-main-sequence binaries also suggest misalignment (e.g., Davis, Mundt & Eislöffel 1994; Roccatagliata et al. 2011). Recently, Jensen & Akeson (2014) showed evidence that at least one of the circumstellar disks in the young binary system HK Tau is inclined with respect to the binary plane by at least 30°. Imaging of circumbinary debris disks indicates that while some are aligned with the binary plane, others can have significant misalignment (such as 99 Herculis, which is misaligned by at least 30°; Kennedy et al. 2012a, 2012b). The pre-main-sequence binary KH 15D is surrounded by a precessing circumbinary disk inclined with respect to the binary plane by 10 – 20° (Chiang & Murray-Clay 2004; Winn et al. 2004; Capelo et al. 2012), and in the FS Tau

This chapter is adapted from Miranda & Lai (2015).

system, circumstellar disks appear to be misaligned with a circumbinary disk (Hioki et al. 2011). Although circumbinary disks may align with the binary plane on relatively short timescales, circumstellar disks are expected to maintain misalignment over timescales comparable to their lifetimes (Foucart & Lai 2014).

In recent years, many exoplanets have been found in binaries, including both S-type planets (orbiting a single member of the binary) using radial velocity measurements (e.g., Hatzes et al. 2003; Chauvin et al. 2011; Dumusque et al. 2012), and P-type (circumbinary) planets using the transit method (e.g., Doyle et al. 2011; Orosz et al. 2012; Welsh et al. 2012, 2014; Kostov et al. 2014). While the orbits of most of the circumbinary planets are consistent with alignment with the binary orbit, Kepler-413b appears to be slightly misaligned (by about 2.5°), possibly as a result of forming in a misaligned circumbinary disk (Kostov et al. 2014).

The sizes of the protoplanetary disks (the outer radius of a circumstellar disk and the inner radius of a circumbinary disk) in which planets in binaries form are affected by gravitational interactions between the disk and binary. The size of the disk is of interest because it places restrictions on the process of planet formation. For example, the process of giant planet formation may be jeopardized in circumstellar disks with significantly truncated outer radii (Jang-Condell 2015).

The size of a pressureless circumstellar disk is determined by the "static", non-resonant tidal force from the binary companion. In particular, the disk size is determined by the location at which orbits around the primary star begin to intersect one another, in the context of the circular restricted three-body problem

(Paczyński 1977; Paczyński & Rudak 1980). In this framework, the outer edge of a circumstellar disk cannot extend past the 2:1 commensurability (i.e., the period ratio of the binary and a test mass in the disk must exceed two), except when the binary mass ratio is extreme. Pichardo, Sparke & Aguilar (2005) extended this analysis to eccentric coplanar binaries, and found that the disk must always be smaller than 73% of the average Roche lobe radius. However, Papaloizou & Pringle (1977) demonstrated that even a very small viscosity can prevent orbit crossings and allow the disk to be larger than the radius determined by non-intersecting orbits.

In addition to static tidal interactions, gas disks (both circumstellar and circumbinary) in binaries are also subject to resonant tidal forcings. The motion of the binary excites spiral density waves at Lindblad resonances, where the natural epicyclic frequency κ of the disk is commensurate with the binary orbital frequency Ω_B. As a result, a torque is exerted on the disk at each resonance (Goldreich & Tremaine 1979, 1980). Artymowicz & Lubow (1994; hereafter AL94) applied the formalism of Goldreich & Tremaine to determine the sizes of circumstellar and circumbinary disks in coplanar binaries. In their approach, the gravitational potential of an eccentric binary is decomposed into many Fourier components, each applying a torque on the disk at Lindblad resonances. The possibility of gap opening and disk truncation by each potential component is determined by the balance between the Lindblad torque and the viscous torque. The main goal of this paper is to generalize the results of AL94 to circumstellar and circumbinary disks in binaries with arbitrary mutual inclinations.

The behavior of Lindblad torques in misaligned disks has been explored in several recent works. Lubow, Martin & Nixon (2015) investigated the torques

experienced by a circumstellar disk in an inclined circular binary, which is a special case of the general binary eccentricities studied in this paper. Nixon & Lubow (2015) demonstrated that an eccentric binary can exert non-zero Lindblad torques on a retrograde circumbinary disk, which are significantly weaker than those applied to a prograde disk, but can nonetheless be responsible for clearing a cavity around the binary. This is a special case of the arbitrary inclination framework used in this paper, although they did not make the assumption of a Keplerian disk. We do not specifically address retrograde disks in this paper.

The most important technical aspect of our work is the computation of the Fourier components of the binary potential for arbitrary inclinations. We represent the potential semi-analytically as a power series in eccentricity (as in the appendix of AL94), evaluated to order e^{10}. This is accomplished through the use of Wigner matrices to account for arbitrary inclination between the disk and binary planes, and through an expansion of the binary orbit using the Hansen coefficients of celestial mechanics. We note that Nixon & Lubow (2015) utilized an exact, non-series approach for computing the gravitational potential for the special case of retrograde circumbinary disks. This method amounts to numerically evaluating the Hansen integrals.

This paper is organized as follows. In Section 5.2 we decompose the gravitational potential of a misaligned binary into Fourier components with various forcing frequencies. In Section 5.3 we review how these potential components exert Lindblad torques on a disk. In Section 5.4 we apply this theory to the truncation of circumstellar disks, determining the size of the disk outer edge. We then determine the inner radii of circumbinary disks in Section 5.5. We address

the validity of our Keplerian disk approximation in Section 5.6. We summarize and discuss our results in Section 5.7.

5.2 Potential Components

We consider a binary consisting of masses M_1 and M_2 with total mass $M_{\rm tot} = M_1 + M_2$, mass ratio $q = M_2/M_{\rm tot}$, semi-major axis a and eccentricity e. The orbital frequency is $\Omega_{\rm B} = \left(GM_{\rm tot}/a^3\right)^{1/2}$. The orbit is inclined relative to a reference plane, taken to be the plane of a thin circumstellar or circumbinary disk, by an angle i. We will express the disturbing potential (per unit mass) acting on the disk in the form

$$\begin{aligned}\Phi &= \sum_{m,\mu,n} \Phi_{m,\mu,n}(r) \cos\left[m\phi - (\mu+n)\,\Omega_{\rm B} t\right] \\ &= \sum_{m,N} \Phi_{m,N}(r) \cos\left(m\phi - N\Omega_{\rm B} t\right),\end{aligned} \tag{5.1}$$

where (r, ϕ) specifies the radial and azimuthal position of the disk particle, m is the azimuthal number in the disk plane, μ is the azimuthal number in the binary plane, N is a time harmonic number, and n is related to the eccentricity dependence of each potential component (see later in this section). In our formulation, $m > 0$, but μ and n can be positive, negative, or zero. Since different (μ, n)-components with the same ϕ and t dependence are not physically distinct, we have defined

$$\Phi_{m,N}(r) = \sum_{\mu,n} \delta_{\mu+n,N} \Phi_{m,\mu,n}(r), \tag{5.2}$$

which measures the total strength of the potential component having the azimuthal number m and rotating with the pattern frequency

$$\omega_{\rm P} = \frac{N}{m}\Omega_{\rm B}. \tag{5.3}$$

Each component, denoted by (m, N), excites density waves at resonant locations in the disk, which give rise to the torques, $T_{m,N}$. We use different approaches to compute the potential $\Phi_{m,N}$, depending on whether we are considering a circumstellar or circumbinary disk.

5.2.1 Circumstellar Disk

We work in the reference frame centered on the primary star M_1, so that the disk rotation rate around M_1 is $\Omega(r) = \left(GM_1/r^3\right)^{1/2}$. The gravitational potential felt by a disk particle due to M_2 is

$$\Phi = -\frac{GM_2}{|\mathbf{r} - \mathbf{r}_2|} + GM_2 \frac{\mathbf{r} \cdot \mathbf{r}_2}{r_{12}^3}, \tag{5.4}$$

where $\mathbf{r}$ is the particle's position vector, and $\mathbf{r}_2 = r_{12}\hat{\mathbf{r}}_2$ is the position vector of M_2 relative to M_1. The second term in the potential is the indirect part arising from the motion of M_1 around the center of mass of the system (e.g., Murray & Dermott 1999). The direct term can be expanded in Legendre polynomials, leading to

$$\Phi = -GM_2 \sum_{l=2}^{\infty} \frac{r^l}{r_{12}^{l+1}} P_l\left(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_2\right). \tag{5.5}$$

The $l = 0$ term is an irrelevant constant which has been dropped. The angular dependence can be separated by an expansion in spherical harmonics,

$$P_l\left(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_2\right) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{l,m}^*\left(\theta_2, \phi_2\right) Y_{l,m}\left(\theta, \phi\right). \tag{5.6}$$

We define two coordinate systems. The unprimed coordinate system (θ, ϕ) has its z-axis aligned with the disk angular momentum, $\hat{\mathbf{z}} = \hat{\mathbf{L}}_{\mathrm{D}}$. The primed coordinate system (θ', ϕ') has its z-axis aligned with the binary angular momentum,

$\hat{\mathbf{z}}' = \hat{\mathbf{L}}_{\rm B}$. The two coordinate systems share a y-axis, and $\hat{\mathbf{z}} \cdot \hat{\mathbf{z}}' = \cos(i)$. The spherical harmonics in these two coordinate systems are related through the Wigner d matrices,

$$Y_{l,m}(\theta_2, \phi_2) = \sum_{\mu=-l}^{l} d^l_{\mu,m}(i)\, Y_{l,\mu}(\theta_2', \phi_2')\,. \tag{5.7}$$

Combining equations (5.5), (5.6) and (5.7), evaluating $Y_{l,m}$ in the disk plane ($\theta = \pi/2$) and $Y_{l,\mu}$ in the binary plane ($\theta' = \pi/2$), and taking only the real part, the potential (5.5) can be written as

$$\begin{aligned}\Phi = -\frac{2GM_2}{a} \sum_{l=2}^{\infty} \sum_{m=1}^{l} \sum_{\mu=-l}^{l} W_{l,m} W_{l,\mu} d^l_{\mu,m}(i) \\ \times \left(\frac{r}{a}\right)^l \left(\frac{r_{12}}{a}\right)^{-(l+1)} \cos(m\phi - \mu\phi_2')\,,\end{aligned} \tag{5.8}$$

where

$$\begin{aligned}W_{l,m} &= \left[\frac{(l-m)!}{(l+m)!}\right]^{\frac{1}{2}} P_l^m(0) \\ &= (-1)^{\frac{l+m}{2}} \left[(l-m)!\,(l+m)!\right]^{\frac{1}{2}} \\ &\times \left[2^l \left(\frac{l-m}{2}\right)! \left(\frac{l+m}{2}\right)!\right]^{-1},\end{aligned} \tag{5.9}$$

with the factor $(-1)^{(l+m)/2}$ taken to be zero if $l + m$ is odd, and similarly for $W_{l,\mu}$. The product $W_{l,m} W_{l,\mu}$ is zero unless l, m and μ are all even or all odd. Notice that we have taken twice the sum over only positive values of m, since the $(-m, -\mu)$ and (m, μ) terms are identical (those with positive or negative μ are still distinct). In this form, the expression for the potential is exact, given an explicit forms of r_{12}/a and ϕ_2', the radial coordinate and true anomaly of M_2 in the binary plane. As a final step, we use their elliptic expansions to write (see Appendix C)

$$\begin{aligned}\left(\frac{r_{12}}{a}\right)^{-(l+1)} \cos(m\phi - \mu\phi_2') = \\ \sum_{n=-\infty}^{\infty} C^{\rm CS}_{l,\mu,n} \cos\left[m\phi - (\mu + n)\,\Omega_{\rm B} t\right].\end{aligned} \tag{5.10}$$

Each coefficient $C^{\rm CS}_{l,\mu,n}$ is a series in powers of e, with the leading term proportional to $e^{|n|}$. The main approximation we make is to truncate these coefficients at

a finite order in eccentricity (e^{10}). The final expression for the potential strengths is

$$\Phi_{m,\mu,n} = -\frac{2GM_2}{a} \sum_{l=l_{\min}}^{\infty} W_{l,m} W_{l,\mu} C^{\mathrm{CS}}_{l,\mu,n} d^l_{\mu,m}(i) \left(\frac{r}{a}\right)^l, \tag{5.11}$$

where $l_{\min} = \max(m, |\mu|, 2)$.

5.2.2 Circumbinary Disk

For circumbinary disks, we work in the reference frame centered on the center of mass of the binary. In this frame, the orbital frequency of a particle in the disk is $\Omega(r) = \left(GM_{\mathrm{tot}}/r^3\right)^{1/2}$. The disturbing potential can then be expressed as

$$\Phi = -\sum_{l=2}^{\infty} \frac{GM_l}{a} \left(\frac{r_{12}}{a}\right)^l \left(\frac{r}{a}\right)^{-(l+1)} P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_2), \tag{5.12}$$

where r_{12} is the separation between M_1 and M_2, and

$$M_l = \left[q(1-q)^l + (-1)^l (1-q) q^l\right] M_{\mathrm{tot}} \tag{5.13}$$

(e.g., Ford, Kozinsky & Rasio 2000; Harrington 1968). In general, odd-l components are weaker than even-l components for similar mass binaries, and for equal mass binaries ($q = 1/2$), M_l is identically zero for odd l. Again writing the Legendre polynomials in terms of spherical harmonics in the binary and disk frames using the Wigner functions [equations (5.6) and (5.7)], we have

$$\begin{aligned} \Phi = -2 \sum_{l=2}^{\infty} \frac{GM_l}{a} \sum_{m=1}^{l} \sum_{\mu=-l}^{l} & W_{l,m} W_{l,\mu} d^l_{\mu,m}(i) \\ & \times \left(\frac{r}{a}\right)^{-(l+1)} \left(\frac{r_{12}}{a}\right)^l \cos(m\phi - \mu\phi'_2). \end{aligned} \tag{5.14}$$

We expand the time-dependent orbital coordinates of the binary in a manner analogous to the circumstellar disk case,

$$\left(\frac{r_{12}}{a}\right)^{l} \cos\left(m\phi - \mu\phi_2'\right) = \sum_{n=-\infty}^{\infty} C^{\mathrm{CB}}_{l,\mu,n} \cos\left[m\phi - (\mu + n)\,\Omega_{\mathrm{B}} t\right], \tag{5.15}$$

so the potential component is given by

$$\Phi_{m,\mu,n} = -2 \sum_{l=l_{\min}}^{\infty} \frac{GM_l}{a} W_{l,m} W_{l,\mu} C^{\mathrm{CB}}_{l,\mu,n} d^{l}_{\mu,m}(i) \left(\frac{r}{a}\right)^{-(l+1)}. \tag{5.16}$$

See Appendix C for the values of the $C^{\mathrm{CB}}_{l,\mu,n}$ coefficients.

5.3 Lindblad Torques

Each potential component $\Phi_{m,N}$, rotating with the pattern frequency $\omega_{\mathrm{P}} = N\Omega_{\mathrm{B}}/m$, excites density waves at the Lindblad resonances (LRs), where

$$\omega_{\mathrm{P}} - \Omega(r) = \pm\frac{\kappa(r)}{m}, \tag{5.17}$$

where the upper (lower) sign corresponds to the outer (inner) LR. From here on we assume that the epicyclic frequency $\kappa(r)$ is equal to $\Omega(r)$ and both are proportional to $r^{-3/2}$, i.e., the disk is exactly Keplerian. The LRs correspond to locations where

$$\frac{\Omega(r)}{\Omega_{\mathrm{B}}} = \frac{N}{m \pm 1}, \tag{5.18}$$

and are located at

$$\frac{r_{\mathrm{LR}}}{a} = \begin{cases} \left[(m \pm 1)/N\right]^{2/3} (1-q)^{1/3} & \text{circumstellar disk} \\ \left[(m \pm 1)/N\right]^{2/3} & \text{circumbinary disk} \end{cases}. \tag{5.19}$$

The torque on the disk at a LR is (Goldreich & Tremaine 1979)

$$T^{\rm LR}_{m,N} = -m\pi^2 \left[\Sigma \left(\frac{{\rm d}D}{{\rm d}\ln r} \right)^{-1} |\Psi_{m,N}|^2 \right]_{r_{\rm LR}}, \tag{5.20}$$

where Σ is the disk surface density, $D = \kappa^2 - m^2 (\Omega - \omega_{\rm P})^2$, and

$$\Psi_{m,N} = \frac{{\rm d}\Phi_{m,N}}{{\rm d}\ln r} + \frac{2\Omega}{\Omega - \omega_{\rm P}} \Phi_{m,N}. \tag{5.21}$$

We also define $\Psi_{m,\mu,n}$, which is the same as the above expression but with $\Phi_{m,\mu,n}$ replacing $\Phi_{m,N}$. While $\Psi_{m,N}$ is the quantity which determines the torque, it is useful to discuss how various components $\Psi_{m,\mu,n}$ contribute to it, since these are physically associated with different azimuthal forcing components of the binary orbit. Note that since we assume Keplerian disks, $2\Omega/(\Omega - \omega_{\rm P}) = \mp 2m$ at $r = r_{\rm LR}$, and

$$\left(\frac{{\rm d}D}{{\rm d}\ln r} \right)_{r_{\rm LR}} = \mp \frac{3N^2}{m \pm 1} \Omega_{\rm B}^2. \tag{5.22}$$

At inner Lindblad resonances (ILRs), $T^{\rm LR}_{m,N} < 0$, i.e., disk particles lose angular momentum, and at outer Lindblad resonances (OLRs), $T^{\rm LR}_{m,N} > 0$, so disk particles gain angular momentum. Torques are also applied to the disk at corotation resonances, where $\omega_{\rm P} - \Omega(r) = 0$, however these are not important in disk truncation. Therefore we subsequently drop the superscript "LR" on $T_{m,N}$, as we only consider Lindblad torques.

The viscous torque on the disk, assuming the α-ansatz for the kinematic viscosity coefficient, $\nu = \alpha c_{\rm s}^2/\Omega$, is given by (e.g., Pringle 1981)

$$T_\nu = 3\pi\alpha h^2 \Sigma \Omega^2 r^4, \tag{5.23}$$

where $h = H/r$ is the disk aspect ratio. As in AL94, we assume a gap is opened at the (m, N) LR if $|T_{m,N}| \geq |T_\nu|$. Throughout this paper, unless otherwise noted, we adopt a disk model with $h = 0.05$ and $\alpha = 0.01$.

5.4 Results: Circumstellar Disk

5.4.1 Resonances Relevant to Outer Disk Truncation

We are interested in the largest outer radius $r_{\rm out}$ that a circumstellar disk can have, given the orbital parameters (a, e, and i) of the binary. This amounts to determining the smallest radius at which a particular resonant torque can open a gap. Therefore we should consider the inner Lindblad resonances of the potential components with the largest pattern frequencies that satisfy $|T_{m,N}| \geq |T_\nu|$, with $|T_\nu|$ evaluated at the resonance location. The pattern frequency associated with $\Phi_{m,N}$ is $\omega_{\rm P} = N\Omega_{\rm B}/m$, so we should examine components with small m and positive N. The smallest m-components whose ILRs are in the disk have $m = 2$, since all $m = 1$ ILRs are formally located at the origin [see equation (5.19)]. Therefore we focus on the $(m, N) = (2, N)$ ILRs with $N \geq 2$, which are located at the $\Omega/\Omega_{\rm B} = N$:1 commensurabilities.

5.4.2 Effects of Disk Inclination

An example of the different contributions to a potential component is depicted in Figure 5.1. Here $\Psi_{m,N} = \Psi_{2,2}$ is shown (solid line) for an equal mass binary with $e = 0.5$ as a function of disk-binary inclination i. The potential is the sum of all $\Psi_{2,\mu,n}$'s with $\mu + n = 2$. The dashed lines depict each of the individual components which significantly contribute to $\Psi_{2,2}$. Other components, not shown in Figure 5.1, such as $\Psi_{2,-4,6}$, contribute negligibly (for $e = 0.5$), but can be important at larger e. We include components up to $|n| = 8$ in all subsequent calculations. At small inclinations, $\Psi_{2,2} \approx \Psi_{2,2,0}$, an equality which is exactly true at

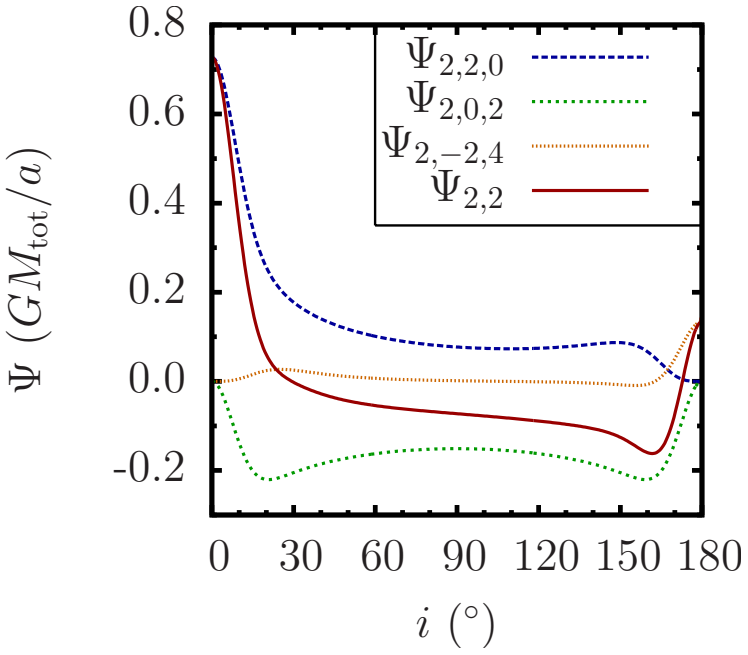

Figure 5.1: The potential component $\Psi_{m,N} = \Psi_{2,2}$ (solid line) and the dominant $\Psi_{2,\mu,n}$ sub-components which contribute to it (dashed lines) for a circumstellar disk in an equal mass binary with $e = 0.5$.

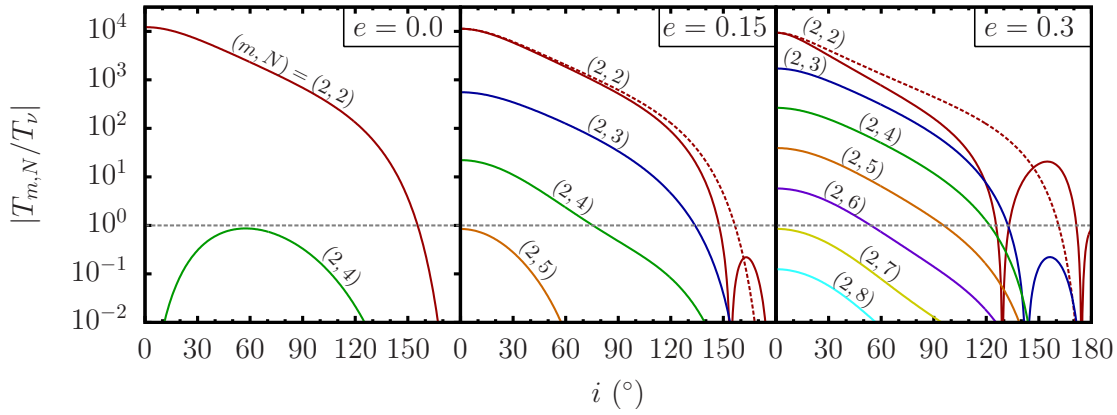

Figure 5.2: Ratio of resonant torque to viscous torque as a function of inclination for the $(m, N) = (2, N)$ inner Lindblad resonances for a circumstellar disk in an equal-mass binary. Each panel corresponds to a different binary eccentricity. The horizontal dashed line corresponds to $|T_{m,N}| = |T_\nu|$, above which a given resonance can clear a gap. The dashed red lines (in the middle and right panels) are the $(2, 2)$ torques computed using only the $\Psi_{2,2,0}$ potential component, rather than a sum of all appropriate $\Psi_{m,\mu,n}$'s.

$i = 0°$, since $d^l_{\mu,m}(0) = \delta_{\mu,m}$. In other words, at small inclinations, the $\mu = 2$ component of the binary potential is primarily responsible for exciting the $m = 2$ disturbances in the disk. Also note that this is the only non-zero component when $e = 0$, since the others have $|n| > 0$, and $\Psi_{m,\mu,n} \propto e^{|n|}$. At intermediate inclinations, the azimuthal components of the binary orbit other than $\mu = 2$ become important, especially the $\mu = 0$ term ($\Psi_{2,0,2}$), which contributes strongly between about 30° and 150°. At inclinations close to 180°, the component $\Psi_{2,-2,4}$ becomes dominant. This component, despite being somewhat suppressed by its eccentricity dependence, becomes significant due to the strong coupling of the $\mu = -2$ and $m = 2$ terms near counter-alignment. Thus, an eccentric counter-aligned binary can exert non-zero Lindblad torques on a circumstellar disk. This is also true for circumbinary disks (see Section 5.5).

The inclination dependence of each component is determined by a weighted sum of Wigner functions, but to leading order is determined by the first term in this sum. For example, $\Psi_{2,2,0} \propto d^2_{2,2}(i) \propto \cos^4(i/2)$ and $\Psi_{2,0,2} \propto d^2_{0,2}(i) \propto \cos^2(i/2)\sin^2(i/2)$, both dominated by the $l = 2$ terms. However, the leading order term is not necessarily a good approximation to the actual i dependence. For example, the leading term of $\Psi_{2,2,0}$ is reduced (compared to $i = 0°$) by 1/4 at 90°, while including all terms shows that this reduction is actually about 1/10, a much steeper decrease with inclination than given by $d^2_{2,2}(i)$. The shape of the total $\Psi_{2,2}$ is therefore a sum of many different Wigner functions, leading to its characteristic shape. There are two unique characteristics of this curve. First, it changes sign twice, at $i = 29°$ and $i = 172°$ (the particular values of these angles depend on the binary eccentricity), so that no torque is exerted on the disk at these inclinations. The sign of $\Psi_{2,2}$ is irrelevant since the torque is proportional to $|\Psi_{2,2}|^2$. Second, properly summing the relevant $\Psi_{m,\mu,n}$'s can counterintuitively

lead to a $\Psi_{m,N}$ whose absolute value (and hence resultant torque) is smaller than $\Psi_{m,m,N-m}$ (the only non-zero component for an aligned disk), as it does for a large range of i in Figure 5.1.

Figure 5.2 shows how the resonant torques responsible for disk truncation (normalized by viscous torque) vary with inclination, for an equal mass binary with several different eccentricities. Notably, even for a circular binary (the left panel of Figure 5.1), there are non-zero torques with $N > 2$ when the disk is not aligned. For example, $|T_{2,4}|$ depends on $\Psi_{2,4}$, which is equal to $\Psi_{2,4,0}$ for a circular binary. To leading order this depends on i as $d^4_{4,2}(i) \propto \cos^6(i/2)\sin^2(i/2)$, which is largest at $i = 60°$, and zero at 0° and 180°. In our canonical disk model (with $h = 0.05$ and $\alpha = 0.01$), $|T_{2,4}|$ is never strong enough to clear a gap when $e = 0$, but if α or h of the disk were slightly smaller, it would be possible for a certain range of inclinations. This is impossible when the disk and binary are aligned. The largest torque for a circular binary is $|T_{2,2}|$, which depends only on $\Psi_{2,2,0}$, a monotonically decreasing function of i (see Figure 5.1). This torque is very large at small inclination (exceeding $|T_\nu|$ by a factor of over 10^4), and gets weaker as inclination increases, becoming about 18 times smaller (compared to aligned) at 90°, and falling to zero at 180°.

For $e > 0$ (the middle and right panels of Figure 5.2), higher N torques generally become stronger (compared to $e = 0$). At small inclinations, $|T_{2,2}| > |T_{2,3}| > |T_{2,4}|$ and so on. This is because at small i these torques are primarily a result of the $\Psi_{2,2,N-2}$ potential components, which have the approximately the same inclination dependence [$d^2_{2,2}(i)$ to leading order], but different eccentricity dependence (e^{N-2}), the latter of which determines their relative strengths. At larger inclinations, the contributions of the $\mu \neq 2$ potential components cause

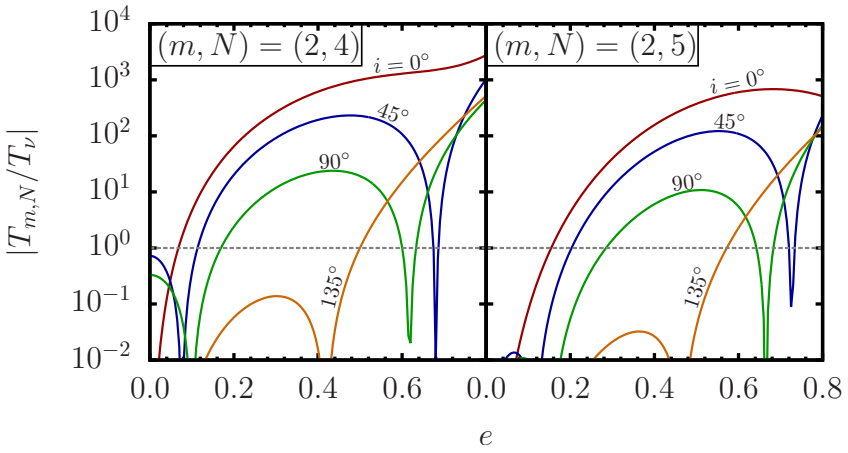

Figure 5.3: Ratio of resonant torque to viscous torque as a function of e for $(m, N) = (2, 4)$ and $(2, 5)$ ILRs in a circumstellar disk, for several inclinations. The horizontal dashed line corresponds to $|T_{m,N}| = |T_\nu|$.

the torques to have more oscillatory inclination dependences (as demonstrated in Figure 5.1). The dashed lines in the middle and right panels of Figure 5.2 show $|T_{2,2}|$ computed using only the $\Psi_{2,2,0}$ component, demonstrating the relative importance of these couplings. As a result of this behavior, at large i (e.g., above about 120° for $e = 0.3$), the torque $|T_{2,N}|$ is no longer monotonic in N.

5.4.3 Location of Outer Disk Edge

The following procedure is used to compute $r_{\rm out}$, the location of the outer edge of a circumstellar disk, for a given disk-binary inclination i. For each resonance (ILR), labeled by (m, N), or by $\Omega(r_{\rm ILR})/\Omega_{\rm B} = N/(m-1)$, we first compute the torque $|T_{m,N}|$ as a function of e (see Figure 5.3). Then we find the range of e for which $|T_{m,N}| > |T_\nu|$ (so that a gap can be opened), for each (m, N). Then at every value of e, the outer radius of the disk is identified as the location of the gap-opening resonance located at the smallest radius.

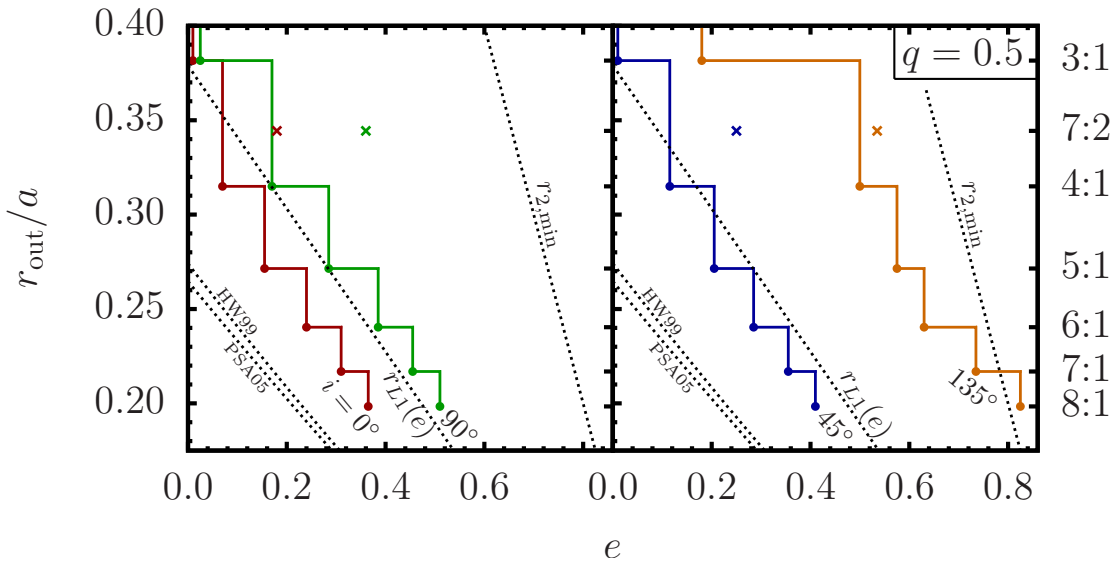

Figure 5.4: Size of a circumstellar disk (location of outer edge $r_{\rm out}$) in an equal mass binary as a function of eccentricity for several disk-binary inclinations. The disk parameters are $h = H/r = 0.05$ and $\alpha = 0.01$. The ratios $\Omega(r_{ILR})/\Omega_{\rm B}$ for the resonances considered are labeled on the right. The filled points indicate the eccentricities and resonance locations at which the outer disk radius changes, and the lines connecting them indicate the behavior of $r_{\rm out}$ between these points. The points represented by crosses indicate the minimum eccentricities required to open a gap at the $\Omega/\Omega_{\rm B} = 7{:}2$ commensurability (for inclinations matched to the lines by color). This is an example of a resonance which is not important for disk truncation, since other resonances can open gaps at smaller radii for the same value of e. From left to right (in each panel), the four dashed lines correspond to: the size of a particle disk determined by Pichardo, Sparke & Aguilar (2005) (labeled "PSA05"), the long-term stability limit for S-type planets in binaries (for $i = 0°$) from Holman & Wiegert (1999) (labeled "HW99"), the average Roche lobe radius evaluated at the pericenter separation of the binary [labeled "$r_{L1}(e)$", see equation (5.24)], and the binary pericenter separation, $r_{2,\rm min} = (1 - e)a$, which sets a strict upper limit for the disk size.

The torques $|T_{m,N}|$ (normalized by viscous torque at the ILR, $|T_\nu|$) for $(m, N) = (2, 4)$ and $(2, 5)$ as a function of e are shown in Figure 5.3. For small e and i, $|T_{m,N}|$ is a monotonically increasing function of e, so that finding the range for which a gap can be opened amounts to finding the minimum e for which $|T_{m,N}| > |T_\nu|$. However, for large inclinations, $|T_{m,N}|$ has local extrema which result in the existence of multiple values of e for which $|T_{m,N}| = |T_\nu|$. For example, at $i = 45°$ and $90°$, as e is increased, $|T_{m,N}/T_\nu|$ exceeds unity above a critical e, then reaches a maximum, decreases, and again drops below unity at a second critical e. Another minimum is then reached, and then it exceeds unity again at yet another critical e. Despite the oscillatory features in Figure 5.3, once the (m, N) resonance is cleared, the $(m, N + 1)$ resonance is always cleared before $|T_{m,N}|$ becomes non-gap-opening. In other words, as e is increased, the innermost gap clearing resonance shifts inward, so that the disk size is a monotonically decreasing function of e for all inclinations we have considered.

Figure 5.4 shows an example of the disk outer radius as a function of eccentricity for an equal mass binary, for several inclinations. In this figure, the filled points represent the eccentricities and resonant locations at which the outer radius abruptly changes due to a new innermost resonance "turning on" (becoming able to clear a gap). These result in the outer disk radius following the stairstep-shaped curves which connect the filled points. Several limiting radii are also shown in Figure 5.4. The dashed line labeled $r_{L1}(e)$ is the average radius of the Roche lobe around M_1 (Eggleton 1983) evaluated at the pericenter of the binary orbit:

$$r_{L1}(e) = \frac{0.49 q_1^{2/3} (1 - e) a}{0.6 q_1^{2/3} + \ln\left(1 + q_1^{1/3}\right)}, \tag{5.24}$$

where $q_1 = M_1/M_2 = (1 - q)/q$ (with $q = M_2/M_{\rm tot}$). Equation (5.24) is approximate since the original Roche lobe is calculated for circular, synchronized binaries.

We see from Figure 5.4 that for $i \gtrsim 90^\circ$, our predicted disk radius (based on gap opening criterion) is larger than $r_{L1}(e)$, suggesting that our result for $r_{\rm out}$ may be larger than the actual disk size. Note, however, that since equation (5.24) approximates the Roche lobe as a sphere, it is possible for the disk to extend beyond $r_{L1}(e)$ for some inclinations. The dashed line labeled $r_{2,\rm min}$ corresponds to the pericenter separation of the binary [$r_{2,\rm min} = (1-e)a$]. This is a strict upper limit on the disk size when it is either aligned or counter-aligned (the disk-projected closest approach distance is slightly different when there is misalignment), since if one of the stars were allowed to plunge through the disk it would rapidly clear material from its orbit.

In Figure 5.4 we have also shown the long-term stability limit for S-type planets in binaries (for $i = 0^\circ$), based on the fitting formula of Holman & Wiegert (1999). This limit is smaller than the truncation radius of the disk, but close to the "orbit-crossing" limit determined by Pichardo, Sparke & Aguilar (2005).

Our canonical disk model has $h = 0.05$ and $\alpha = 0.01$. The effect of varying the disk properties is shown in Figure 5.5, which depicts the disk outer radius as a function of eccentricity for an equal mass binary (as in Figure 5.4) when $\alpha = 10^{-3}$ and $\alpha = 10^{-4}$. These correspond to a reduction of all viscous torques T_ν by a factor of 10 and 100 compared to our canonical model. Since T_ν is proportional to αh^2, the curves labeled $\alpha = 10^{-3}$ represent any disk with a combination of α and h such that $\left(\alpha/10^{-3}\right)(h/0.05)^2 = 1$, and similarly for curves labeled $\alpha = 10^{-4}$. Although the viscous torques rescale trivially in this way, the complicated dependence of the resonant torques on eccentricity and inclination does not allow a simple rescaling of the disk size versus eccentricity relations. This explains the qualitative differences between the curves in Figures 5.4 and 5.5. Most notably,

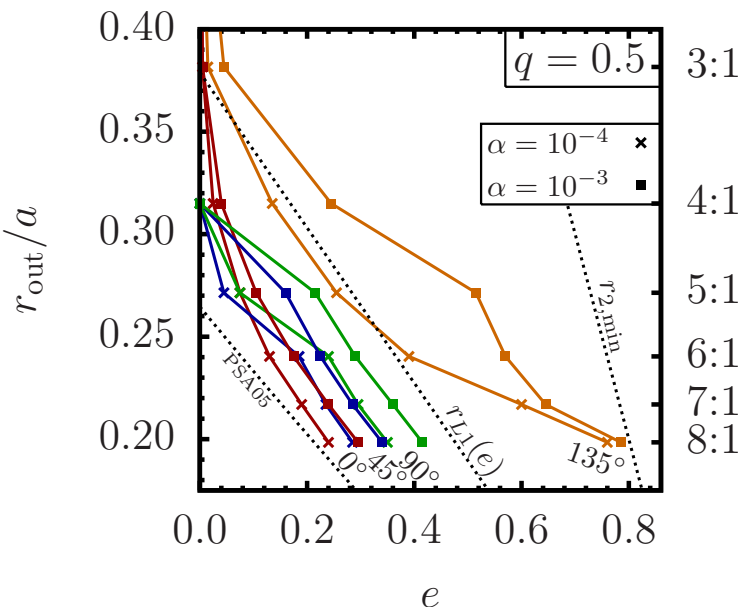

Figure 5.5: Same as Figure 5.4, except for for disks with viscosity parameter $\alpha = 10^{-3}$ and 10^{-4}. The straight lines connecting the points are for graphical convenience only, and are a proxy for the starstep-shaped curves depicted in Figure 5.4.

for the values of α in Figure 5.5, a gap can be opened at the 4:1 commensurability at zero eccentricity for $i = 45°$ and 90°, contrary to our canonical disk model (see Figure 5.4). However, broadly speaking, the difference in disk size due to an order of magnitude change in αh^2 is comparable to the difference in size due to a change of inclination of about 45°.

The effect of binary mass ratio on circumstellar disk size is explored in Figure 5.6. Recall our definition of the mass ratio $q = M_2/M_{\rm tot}$, and that M_2 is always considered the perturber, regardless of whether it is more or less massive than M_1. Therefore the disk sizes for $q = 0.1$ and $q = 0.9$ can be thought of as the sizes of the circumprimary and circumsecondary disks in a system in which the secondary is 1/9 the mass of the primary, and similarly for $q = 0.3$ and $q = 0.7$ (in which case the secondary is 3/7 the mass of the primary). The behavior

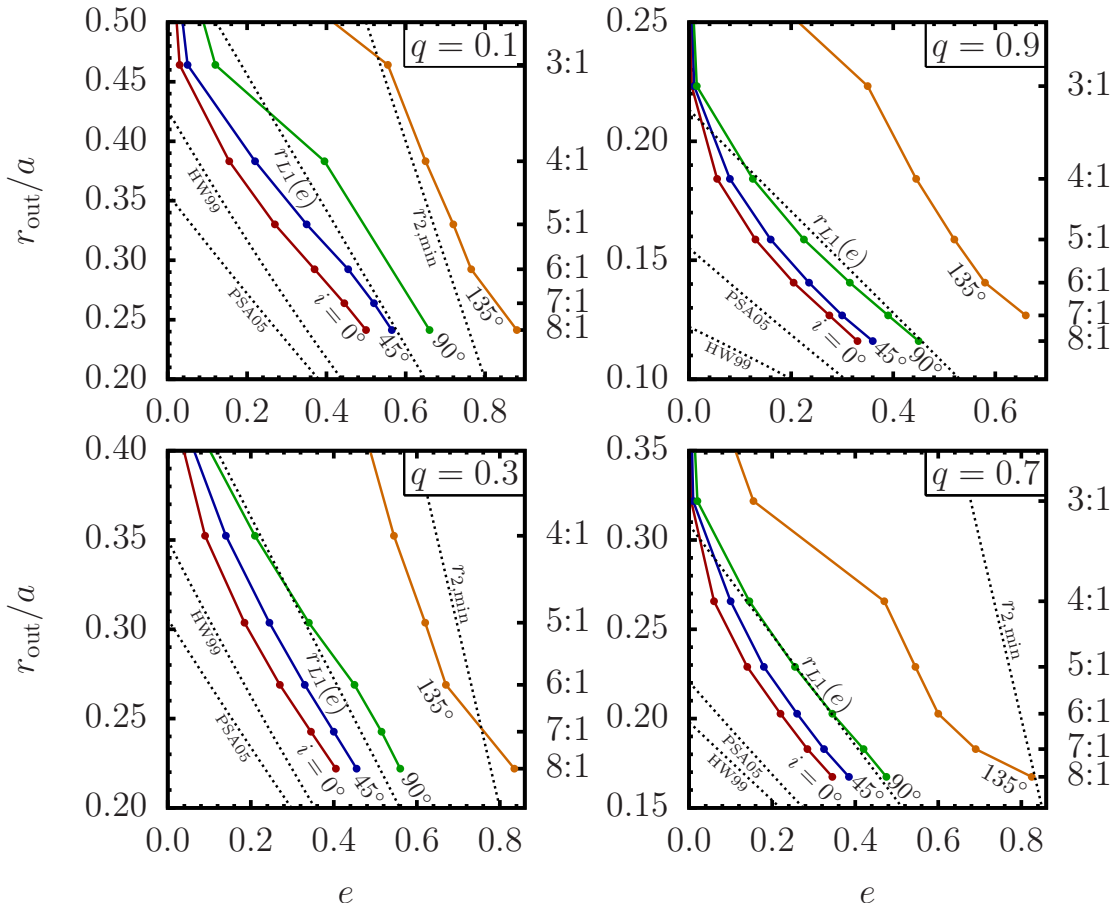

Figure 5.6: Same as Figure 5.4, except for several different mass ratios. As in Figure 5.5, the straight lines connecting the points are only for graphical convenience, and should be interpreted as stairstep-shaped depicted in Figure 5.4.

is qualitatively similar to the equal mass case (see Figure 5.4), but with more massive perturbers leading to smaller disks and vice versa.

We have restricted our calculation of $r_{\rm out}$ to binaries with $e \lesssim 0.8$. At high eccentricity, our ability to compute the disk outer radius is restricted by our finite expansion of the disturbing potential: the highest order resonance we consider is $\Omega/\Omega_{\rm B} = 8{:}1$. For larger values of e, higher order resonances can clear gaps, further reducing the size of the disk, but computing the appropriate potential components for these resonances is impractical in our semi-analytic approach.

5.5 Results: Circumbinary Disk

5.5.1 Resonances Relevant to Inner Disk Truncation

Determining the size $r_{\rm in}$ of the inner cavity of a circumbinary disk amounts to finding the largest radius at which a gap can be cleared. Therefore we consider the OLRs of the potential components with the smallest possible pattern frequencies. Since $\Omega(r_{\rm OLR})/\Omega_{\rm B} = N/(m+1)$ [see equation (5.18)], we shall focus on $N = 1$ and $m \geq 1$, corresponding to the $\Omega/\Omega_{\rm B} = 1{:}(m+1)$ commensurabilities. However, for equal-mass or nearly-equal mass binaries, the strength of the odd-m potential components is zero or small relative to the even-m components, resulting in a relatively weak torque $|T_{m,N}|$. As we will also show, $|T_{1,1}|$ can be weak even for binaries which are not close to equal mass. Therefore, it is also necessary to consider the next strong resonance interior to the 1:2 commensurability, namely the 2:3 commensurability ($m = 2$, $N = 2$), which can be cleared by the $T_{2,2}$ torque.

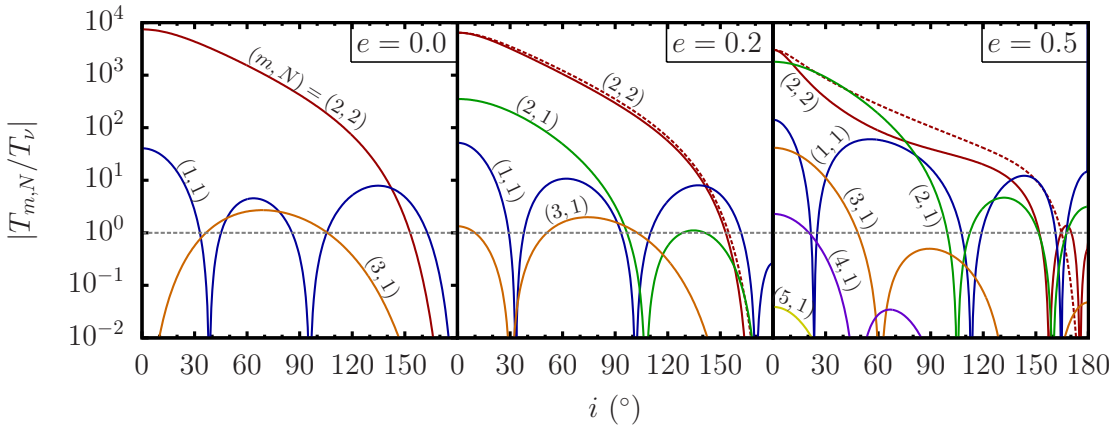

Figure 5.7: Ratio of resonant torque to viscous torque ratio as a function of inclination for the $(m, N) = (m, 1)$ and $(2, 2)$ outer Lindblad resonances in a circumbinary disk with mass ratio $q = 0.3$. Each panel corresponds to a different binary eccentricity. As in Figure 5.2, the horizontal dashed line corresponds to $|T_{m,N}| = |T_\nu|$ (above which a resonance can clear a gap), and the dashed lines in the middle and right panels are the $(2, 2)$ torques computed using only the $\Psi_{2,2,0}$ potential component.

5.5.2 Effects of Disk Inclination

Figure 5.7 shows the resonant torques normalized by viscous torque for a circumbinary disk around a binary with mass ratio $q = 0.3$ (cf. Figure 5.2 for a circumstellar disk). The $(m, N) = (1, 1)$ torque is weak due to the absence of the dipole term in the binary potential [see equation (5.12)], so that to leading order, $\Psi_{1,1}$ is proportional to $d^3_{1,1}(i)(r/a)^{-4}$. Meanwhile $\Psi_{2,2}$ is approximately proportional to $d^2_{2,2}(i)(r/a)^{-3}$ (and has its OLR at a smaller radius), and so the $(m, N) = (2, 2)$ torque is generally much stronger than the $(1, 1)$ torque and can potentially be the most relevant to clearing the inner cavity at low e and i. The $(2, 2)$ component is also relevant for an equal-mass binary, for which all torques with odd m are zero.

For the same reasons as for a circumstellar disk, the torque can be an oscil-

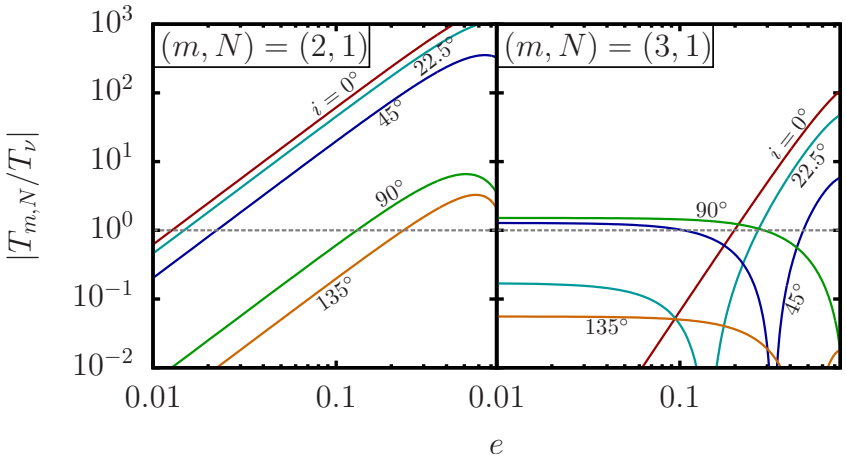

Figure 5.8: Ratio of resonant torque to viscous torque as a function of e for $(m, N) = (2, 1)$ and $(3, 1)$ outer Lindblad resonances in a circumbinary disk, for a mass ratio $q = 0.3$. The dashed horizontal line indicates $|T_{m,N}| = |T_\nu|$.

latory function of i, but this effect is even more pronounced for a circumbinary disk. For example, for a circular binary, $T_{3,1}$ can clear a gap at the 1:4 commensurability for i between 37° to 106°, which is impossible for an aligned disk. For $e > 0$, there are several ranges of i for which the ordering of torques is very different than for an aligned disk, for which the torque strengths depend monotonically on the resonance location (innermost resonances are the strongest).

5.5.3 Inner Cavity Size: Inner Disk Radius

The procedure for computing the inner disk radius $r_{\rm in}$ for a circumbinary disk at a given inclination i is as follows. First, for each resonance $(m, N) = (m, 1)$ and $(2, 2)$, we compute $|T_{m,N}|$ as a function of e (see Figure 5.8), and determine the range of e values where gap opening is possible ($|T_{m,N}| \geq |T_\nu|$). The inner disk radius is determined by which gap-opening resonance is located at the largest radius for every value of e.

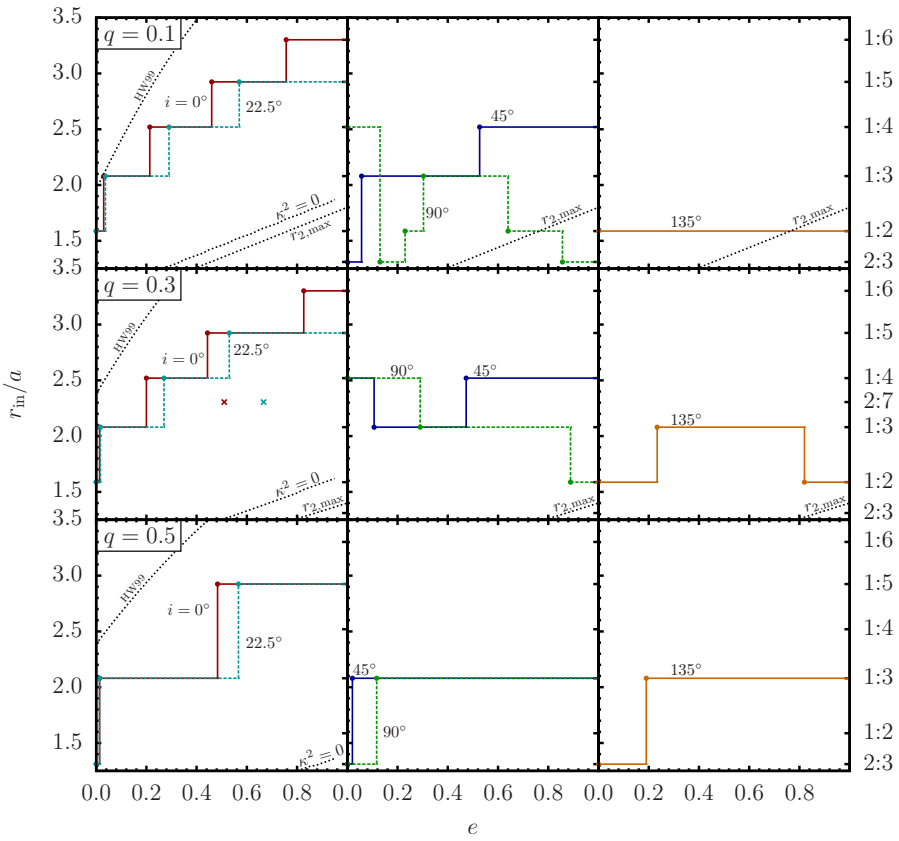

Figure 5.9: Size of inner cavity (location of inner edge $r_{\rm in}$) of a disk around a binary with several mass ratios (corresponding to different rows), each for several binary-disk inclinations. The disk parameters are $h = 0.05$ and $\alpha = 0.01$. The ratios $\Omega(r_{\rm OLR})/\Omega_{\rm B}$ for the resonances considered are labeled on the right. The filled points indicate the eccentricities and resonance locations at which the inner disk radius changes, and the lines connecting them indicate the behavior between these points. In the middle row, the points represented by crosses indicate the minimum eccentricities for which a gap can be cleared at the 2:7 commensurability, for $i = 0°$ and 22.5° (for higher inclinations, a gap cannot be cleared for any value of e). This is an example of a resonance which is not important in determining the size of the inner disk, since gaps can be opened at larger radii for the same values of e. The dashed lines in the left panels labeled "$\kappa^2 = 0$" indicate the radius below which the disk does not satisfy the Rayleigh stability criterion (for $i = 0°$). The dashed lines labeled "$r_{2,\rm max}$" correspond to the radial location of the secondary component of the binary at apocenter [$r_{2,\rm max} = (1 - q)(1 + e)a$], which is a strict lower limit on $r_{\rm in}$ (for $q = 0.5$, $r_{2,\rm max}$ is below the displayed range of the y-axis). The dashed lines labeled "HW99" indicate the stability limit for P-type planets in binaries (for $i = 0°$) from Holman & Wiegert (1999). Note that for $q = 0.3$ and $q = 0.5$, when $i = 0°$ and $i = 22.5°$, the cavity transitions from its smallest size (the 1:2 commensurability for $q = 0.3$ and the 2:3 commensurability for $q = 0.5$) to a larger size (1:3 in both cases) at a very small but non-zero eccentricity ($e = 0.01 - 0.02$).

As in the circumstellar disk case, the complicated dependence of $|T_{m,N}|$ on e can result in several ranges of e for which a resonance can open a gap. But unlike the circumstellar disk case, this effect is significant enough to qualitatively alter the dependence of the disk size on e. As an example, $|T_{m,N}/T_\nu|$ is shown in Figure 5.8 for $(m, N) = (2, 1)$ and $(3, 1)$, corresponding to the $\Omega/\Omega_B = 1{:}3$ and 1:4 commensurabilities, for a binary with mass ratio $q = 0.3$, for several disk inclinations. While $|T_{2,1}|$ is a monotonically increasing function of e, such that for each i there is a critical e above which it can open a gap, the behavior of $|T_{3,1}|$ is very different. For $i = 45°$ and 90°, it can open a gap at zero eccentricity, and it decreases with e so that above a certain critical e a gap can no longer be opened (for $i = 45°$ a gap can again be opened at yet another larger e). In general, a given resonance, (m, N), is not always gap-opening at the lowest value of e for which the next resonance, $(m + 1, N)$, first opens a gap. This is in contrast to the case of a circumstellar disk, for which gaps are always cleared at sequentially higher order resonances as e is increased. At large inclinations, this vastly different behavior of the $|T_{m,N}|$'s can result in cavity sizes which both increase or decrease with increasing e, sometimes exhibiting both behaviors for a single inclination.

The inner disk radii for three different mass ratios and various inclinations are shown in Figure 5.9. We first focus on the middle row, which corresponds to the mass ratio $q = 0.3$ (as in Figures 5.7 and 5.8). For $i = 0°$ and 22.5°, the inner edge is located at the 1:2 commensurability for $e = 0$ and increases with e as gaps are opened at higher-order resonances. The maximum cavity size is located at the 1:6 commensurability $i = 0°$, and at the 1:5 commensurability for $i = 22.5°$. For $i = 45°$ and 90°, the 1:4 commensurability can be cleared at zero eccentricity (see the left panel of Figure 5.7), but the cavity size then decreases with e as the 1:4 torque becomes too small and the 1:3 torque becomes responsible for gap

opening. For $i = 45°$, the size then goes back up to the 1:4 commensurability at $e = 0.47$ and remains at that size for larger e. For $i = 90°$ the cavity shrinks as e increases, then remains at the 1:2 commensurability. Clearly, the behavior of the disk inner cavity size is more complicated than in the circumstellar disk case, and even its qualitative behavior (increasing or decreasing with e) cannot be trivially ascertained, but requires a full consideration of the details of the resonant torques.

The top and bottom rows of Figure 5.9 give the inner disk radii for $q = 0.1$ and $q = 0.5$. Note that a mass ratio q greater than $1/2$ is degenerate with a mass ratio of $1 - q$ for a circumbinary disk. Much of the qualitative behavior seen in the $q = 0.3$ case can also be seen for $q = 0.1$, for example the cavity size is an increasing function of e for $i \leq 45°$, while for $i = 90°$, it can either be decreasing or increasing over different ranges of e. A new behavior is seen for $i = 135°$ and $q = 0.1$: the disk size is indepedendent of e. In this case, for all values of e, the torque $|T_{1,1}|$ is strong enough to clear the 1:2 commensurability, but no other torque $|T_{m,N}|$ can clear a gap at its resonant location in the disk regardless of the value of e. For an equal mass binary ($q = 0.5$), there are fewer possible truncation sites because odd m torques are zero (since $M_l = 0$ for odd l). The only allowed cavity sizes are at the 2:3, 1:3 and 1:5 commensurabilities. The dependence of the cavity size on e is simple in this case: for all inclinations, $r_{\rm in}$ increases with e, with the maximum cavity size located at the 1:5 commensurability for $i = 0°$ and $i = 22.5°$ and at 1:3 commensurability for other inclinations (45°, 90°, 135°).

Figure 5.9 also shows, as a function of eccentricity, the radius at which the squared radial epicyclic frequency [see equation (5.26)] is equal to zero. Below this radius, disk particles are unstable to radial perturbations (according to the

Rayleigh stability criterion), and so the disk can not extend inwards beyond this radius. This radius is shown in the left panels, and is evaluated for $i = 0°$ (or equivalently, $i = 180°$). It is not shown for larger inclinations, for which the instability region shifts inward and lies interior to $r_{2,\max} = (1 - q)(1 + e)a$, the radial coordinate of the secondary star at the apocenter of the binary orbit (also shown in Figure 5.9), which is a strict lower limit for $r_{\rm in}$. For the mass ratios and inclinations we have considered, neither of these non-resonant truncation mechanisms constrain the inner cavity size, except for at very large e for $q = 0.1$ and $i \geq 90°$ (see the two upper right panels), for which the Lindblad torques are weak. Note that for a retrograde disk ($i = 180°$), for which the torques are further weakened, the Rayleigh stability criterion may be more relevant for truncation than gap clearing by Lindblad torques for a wider range of binary parameters (see Nixon & Lubow 2015).

The long-term stability limit for P-type planets in binaries (for i = 0°) (Holman & Wiegert 1999) is also shown in Figure 5.9. The inner truncation radius is inside the stability limit, which is of interest to the formation of observed circumbinary planets, many of which reside very close to the stability limit (see Welsh et al. 2014 and references therein). Thus, planets that have formed near or migrated to the inner edge of the disk (e.g., Kley & Haghighipour 2014) may experience dynamical instability as the disk disappears.

5.6 Effects of Non-Keplerian Rotation

Up to this point, we have considered disks with Keplerian rotation profiles ($\Omega \propto r^{-3/2}$), for which the rotation frequency $\Omega(r)$ is equal to the radial epicyclic

frequency $\kappa(r)$. Strictly, this is valid when the static, axisymmetric part of the gravitational potential consists only of a monopole term, proportional to $1/r$. However, in general, the rotation profile is modified by the quadrupole and higher-order components of the binary potential.[1] The squared rotation frequency is given by

$$\Omega^2(r) = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left[\Phi_{\mathrm{K}}(r) + \Phi_{0,0}(r)\right], \tag{5.25}$$

and the squared radial epicyclic frequency by

$$\kappa^2(r) = \left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{3}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)\left[\Phi_{\mathrm{K}}(r) + \Phi_{0,0}\,(r)\right]. \tag{5.26}$$

Here $\Phi_{\mathrm{K}}(r)$ is equal to $-GM_1/r$ for a circumstellar disk and $-GM_{\mathrm{tot}}/r$ for a circumbinary disk, and $\Phi_{0,0}(r)$ is the $(m, N) = (0, 0)$ component of the appropriate disturbing potential. Equations (5.25) and (5.26) assume that pressure gradients, which are an additional source of non-Keplerian rotation, contribute negligibly compared to gravity (see Section 5.7).

The non-Keplerian rotation profile has two effects on the Lindblad resonances. First, the resonance locations are shifted, so that they no longer correspond to exact integer commensurabilities of $\Omega(r)$ and Ω_{B} [as in equation (5.19)]. Instead, they are determined by solving the general LR condition [equation (5.17)], and are functions of e and i. We adopt the notation $r_{\mathrm{LR}}^{\mathrm{K}}$ for the Keplerian LR locations [as given by equation (5.19)] to distinguish them from exact locations, r_{LR}. Second, the Lindblad torques [equation (5.20)] are modified from their values in a Keplerian disk, since $\mathrm{d}D/\mathrm{d}\ln r$ and $\Psi_{m,N}$ must be evaluated at the new resonance locations using the modified expressions for Ω and κ. We also adopt the notation $T_{m,N}^{\mathrm{K}}$ for the torque on a Keplerian disk to distinguish it from the true torque $T_{m,N}$. In this section, we consider how these effects mod-

[1] Nixon & Lubow (2015) included this effect in their analysis of the Lindblad torques experienced by a retrograde circumbinary disk.

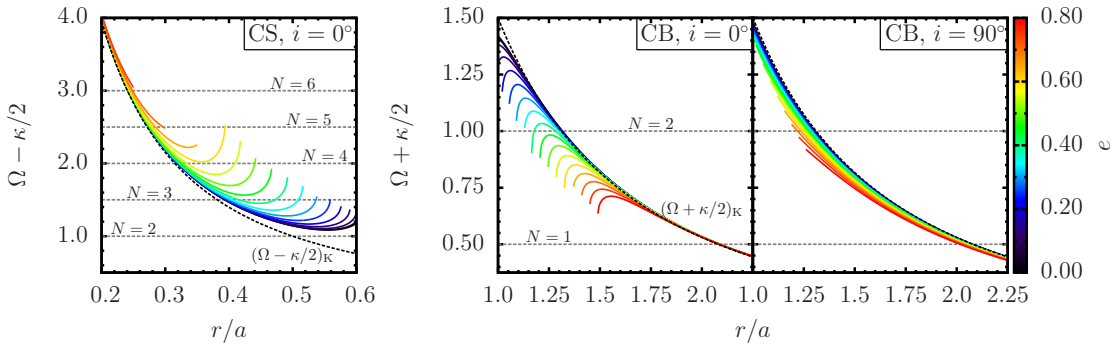

Figure 5.10: The resonant frequencies (combinations of rotation frequency Ω and radial epicyclic frequency κ) which determine the location of several Lindblad resonances [see equation (5.17)]: $\Omega - \kappa/2$ (in units of the binary orbital frequency Ω_B), which is relevant to $m = 2$ ILRs for a circumstellar disk (shown for an equal mass binary with $i = 0°$; left), and $\Omega + \kappa/2$, which is relevant to $m = 2$ OLRs (shown for a circumbinary disk with $q = 0.3$, for $i = 0°$ and $i = 90°$; right). The line colors correspond to different binary eccentricities, as indicated by the color bar on the right. The curves terminate (at large r for a circumstellar disk, and at small r for a circumbinary disk) at the point at which κ^2 becomes negative. The dashed curves indicate the corresponding frequencies ($\Omega \pm \kappa/2$) in a Keplerian disk. The horizontal dashed lines represent pattern frequencies $\omega_P = N\Omega_B/2$ for several values of N. The intersection of the $\Omega \pm \kappa/2$ curves with the horizontal lines correspond to the locations of the $(m, N) = (2, N)$ LRs. Note that for a circumstellar disk, Ω and κ do not strongly deviate from their Keplerian values at large inclinations, therefore only the $i = 0°$ case is shown.

ify the properties of the resonances relevant to truncation for both circumstellar and circumbinary disks.

5.6.1 Circumstellar Disk

Figure 5.10 (left panel) gives an example of how the disk rotation profiles and the locations of ILRs (for $m = 2$) are modified by a non-Keplerian potential. The profiles are obtained by evaluating equations (5.25) - (5.26) numerically. At high eccentricities, the strong deviations from Keplerian rotation render some resonances non-existent, since there is no r for which $\Omega - \kappa/m$ is equal to ω_P. This

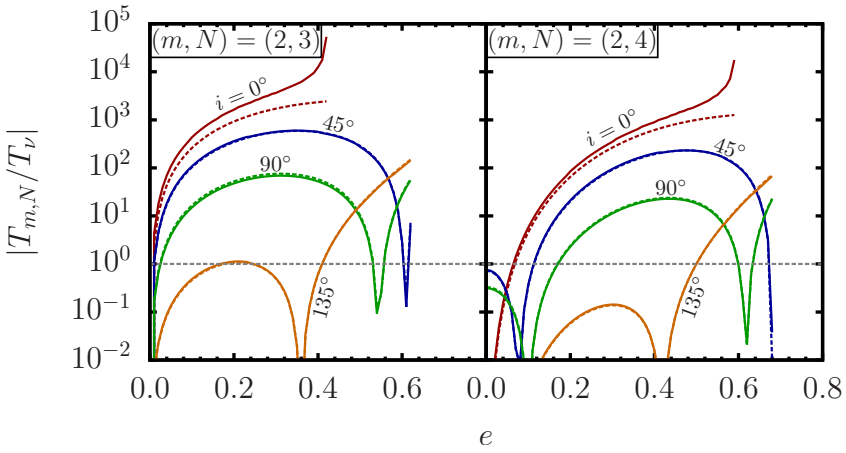

Figure 5.11: Ratio of resonant torque $T_{m,N}$ to viscous torque T_ν, both evaluated at the true resonance location $r_{\rm LR}$ (solid lines), and the ratio of these torques in a Keplerian disk, evaluated at the Keplerian resonance location $r_{\rm LR}^{\rm K}$ (dashed lines), for a circumstellar disk in an equal mass binary. The $(m, N) = (2, 3)$ and $(2, 4)$ ILRs (left and right panels), which are the lowest-order resonances responsible for truncation at non-zero eccentricity, are each shown for several inclinations. The horizontal dashed line corresponds to the threshold for disk truncation ($|T_{m,N}/T_\nu| = 1$).

occurs for $e \gtrsim 0.4$ for the $N = 3$ ILR, and for $e \gtrsim 0.6$ for the $N = 4$ ILR. Notably, the $(m, N) = (2, 2)$ ILR (nominally the 2:1 orbital commensurability) does not exist for an aligned disk in a circular, equal mass binary. The largest fractional shift in $r_{\rm LR}$ for a given resonance is about 12%, occurring for the largest value of e for which the resonance still exists. For eccentricities near this value, there can also be a second location satisfying the resonant criterion. However, the second location is unlikely to be relevant to disk truncation, as it is close to the point at which the disk becomes Rayleigh unstable, and is located at a larger r than the original resonance.

The numerical results for the ratios of resonant torque $T_{m,N}$ to viscous torque T_ν are depicted in Figure 5.11, along with their Keplerian counterparts. Note that the correction to T_ν at the modified resonance location has been taken into

account in Figure 5.11. The deviation of the torque ratio from its Keplerian counterpart is largest for $i = 0°$, and negligible for other inclinations (comparable to the thickness of the plotted lines). The $i = 0°$ curves terminate at a smaller value of e than for other inclinations, due to the fact that the resonance ceases to exist (see Figure 5.10). For eccentricities close to this point, the torque deviates strongly from its Keplerian value (by over an order of magnitue), due to the rotation profile becoming highly distorted. However, these large deviations occur at much larger values of e than the ones for which the resonance first truncates the disk. The eccentricities for which $|T_{m,N}| = |T_\nu|$ change by less than 0.01 for the resonances shown in Figure 5.11. Higher order resonances, located at smaller radii, are affected even less strongly. Therefore, the assumption of a Keplerian disk is a reasonable approximation for determining outer disk truncation.

To quadrupole ($l = 2$) and e^2 order, explicit expressions for the disk rotation profiles and Lindblad torques can be obtained. The axisymmetric, time independent part of the disturbing potential can be approximated by

$$\Phi_{0,0}(r) \approx -\frac{GM_2}{4a} f_{\rm CS}(e,i) \left(\frac{r}{a}\right)^2, \tag{5.27}$$

where the dependence on the inclination and eccentricity of the binary is given by

$$f_{\rm CS}(e,i) = \frac{1}{2}\left[3\cos^2(i) - 1\right]\left(1 + \frac{3}{2}e^2\right). \tag{5.28}$$

Thus, to this order,

$$\Omega^2(r) \approx \Omega_{\rm K}^2 + \frac{2\Phi_{0,0}(r)}{r^2} \quad \text{and} \quad \kappa^2(r) \approx \Omega_{\rm K}^2 + \frac{8\Phi_{0,0}(r)}{r^2}. \tag{5.29}$$

The resonance condition for $m = 2$ ILRs, which are relevant to outer disk truncation, is then

$$\Omega_{\rm K}\left(1 + 2\epsilon_{\rm CS}\right) \approx N\Omega_{\rm B}, \tag{5.30}$$

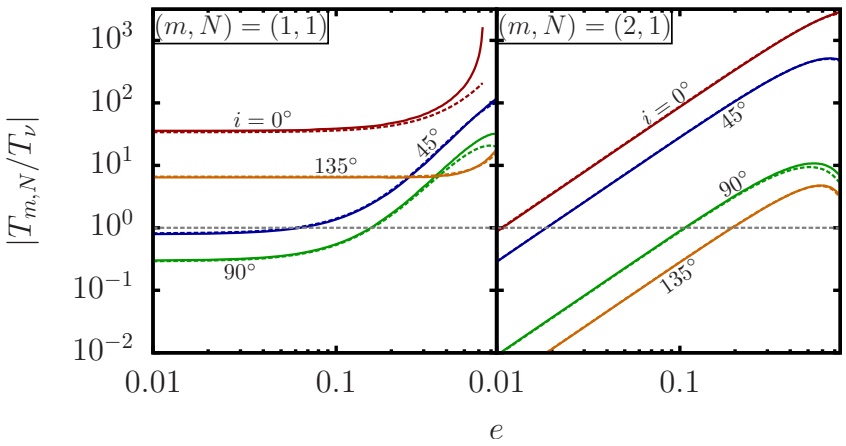

Figure 5.12: Same as Figure 5.11, but for the $(m, N) = (1, 1)$ and $(2, 1)$ OLRs in a circumbinary disk with mass ratio $q = 0.3$.

where we have defined the dimensionless parameter

$$\epsilon_{\rm CS} = \left(\frac{\Phi_{0,0}}{\Phi_{\rm K}}\right)_{r^{\rm K}_{\rm LR}} = \frac{q f_{\rm CS}(e, i)}{4N^2}. \tag{5.31}$$

The resonances are therefore located at

$$r_{\rm LR} \approx r^{\rm K}_{\rm LR}\left(1 + \frac{4}{3}\epsilon_{\rm CS}\right), \tag{5.32}$$

and hence shift outward (inward) compared to their locations in a Keplerian disk when i is less (greater) than 54.7°. We can use equations (5.29) and (5.32) to evaluate the Lindblad torque $T_{m,N}$ [equation (5.20)] explicitly (to quadrupole and e^2 order). For $m = 2$, we find

$$T_{2,N}(r_{\rm LR}) \approx T^{\rm K}_{2,N}\left(r^{\rm K}_{\rm LR}\right)\left(1 + \frac{58}{3}\epsilon_{\rm CS}\right). \tag{5.33}$$

5.6.2 Circumbinary Disk

In Figure 5.10 (middle and right panels), we show the locations of the $(m, N) = (2, 1)$ and $(2, 2)$ OLRs for $i = 0°$ and 90°, including the effects of the non-Keplerian

potential. The largest fractional shifts are about 4%, provided that the resonance exists. The $(m, N) = (2, 2)$ resonance does not exist for $e \gtrsim 0.5$ when $i = 0°$, and for $e \gtrsim 0.7$ when $i = 90°$.

The ratio of resonant to viscous torques is shown in Figure 5.12. The $(m, N) = (1, 1)$ torque differs most from the Keplerian case for $i = 0°$, especially near $e \approx 0.6$, above which the resonance no longer exists. For $i = 90°$, there is also some appreciable deviation from the Keplerian torque at large values of e. However, the values of e for which the resonances become gap-opening change negligibly. Therefore, as for the case of a circumstellar disk, the approximation of a Keplerian disk does not affect truncation.

To leading order in l and e, the non-Keplerian part of the potential can be approximated by

$$\Phi_{0,0} \approx -\frac{G\mu_{\rm B}}{4a} f_{\rm CB}(e, i) \left(\frac{r}{a}\right)^{-3}, \tag{5.34}$$

with

$$f_{\rm CB}(e, i) = \frac{1}{2}\left\{\left[3\cos^2(i) - 1\right]\left(1 + \frac{3}{2}e^2\right) - 15e^2\sin^2(i)\right\}, \tag{5.35}$$

where $\mu_{\rm B} = M_1 M_2 / M_{\rm tot}$ is the reduced mass of the binary. The resulting rotation frequency and epicyclic frequency are

$$\Omega^2(r) \approx \Omega_{\rm K}^2 - \frac{3\Phi_{0,0}(r)}{r^2} \quad \text{and} \quad \kappa^2(r) \approx \Omega_{\rm K}^2 + \frac{3\Phi_{0,0}(r)}{r^2}. \tag{5.36}$$

The resonance condition for $N = 1$ OLRs, which are relevant to inner disk truncation, is

$$\Omega_{\rm K}\left[1 + \frac{3}{2}\left(\frac{m-1}{m+1}\right)\epsilon_{\rm CB}\right] \approx \frac{\Omega_{\rm B}}{m+1}, \tag{5.37}$$

where we have defined the dimensionless parameter

$$\epsilon_{\rm CB} = \left(\frac{\Phi_{0,0}}{\Phi_{\rm K}}\right)_{r_{\rm LR}^{\rm K}} = \frac{q(1-q) f_{\rm CB}(e, i)}{4(m+1)^{4/3}}. \tag{5.38}$$

So the resonance locations are

$$r_{\rm LR} \approx r_{\rm LR}^{\rm K}\left[1+\left(\frac{m-1}{m+1}\right)\epsilon_{\rm CB}\right]. \tag{5.39}$$

In this approximation, the $(m,N)=(1,1)$ resonance (nominally the $\Omega/\Omega_{\rm B}=1{:}2$ commensurability) is not shifted relative to its position in a Keplerian disk. The other $N=1$ resonances shift to larger (smaller) radii compared to their locations in a Keplerian disk when $f_{\rm CB}(e,i)$ is positive (negative). Explicitly evaluating the Lindblad torque to the same order of approximation, we find that for $N=1$ (and $m>1$),

$$\begin{aligned} T_{m,1}\left(r_{\rm LR}\right) \approx\ & T_{m,1}^{\rm (K)}\left(r_{\rm LR}^{\rm K}\right) \\ & \times\left\{1+\left[\frac{m+2}{m+1}+\frac{12m}{3m+1}-2(m-1)\right]\epsilon_{\rm CB}\right\}. \end{aligned} \tag{5.40}$$

5.7 Summary and Discussion

5.7.1 Main Results

We have developed a method for computing Lindblad torques due to a binary potential on circumstellar and circumbinary disks which are misaligned with the binary orbital plane. We used this theory to determine the outer radii of circumstellar disks and the inner radii of circumbinary disks, generalizing the work of Artymowicz & Lubow (1994; AL94) for aligned disks. The summary of our results is as follows.

In the presence of misalignment (and non-zero eccentricity), each azimuthal component of the binary potential experienced by the disk is a result of many azimuthal components of the potential in the binary plane. This is in contrast to aligned disk, for which each azimuthal component of the potential can only

produce perturbations in the disk which have the same azimuthal number. As a result, the inclination and eccentricity dependence of the gap-opening Lindblad torques can be somewhat complicated. Rather than simply decreasing with inclination i (as for circular binaries), or increasing with eccentricity e (as for aligned disks), the Lindblad torque, $|T_{m,N}|$, associated with the potential component which rotates with pattern frequency $\omega_{\rm P} = N\Omega_{\rm B}/m$ (where $\Omega_{\rm B}$ is the binary orbital frequency, and m, N are integers), generally has multiple extrema as functions of e and i.

As in AL94, we adopt the resonance gap opening criterion $|T_{m,N}| > |T_\nu|$, where T_ν is the viscous torque evaluated at the resonance location. For circumstellar disks, the most important resonances for disk truncation are located at the $\Omega/\Omega_{\rm B} = N$:1 commensurabilities, with $N \geq 2$. Ignoring non-resonant truncation mechanisms, the outer edge of the disk is determined by the innermost of these resonances which can clear a gap. Despite the complicated dependence of inner Lindblad torques $|T_{m,N}|$ on inclination and eccentricity, for a given inclination, the resultant outer disk radius is a decreasing function of eccentricity (see Figures 5.4–5.6). This is the same qualitative behavior as for an aligned disk (AL94). Larger inclinations lead to larger disks: in an equal mass binary, the disk is about 20% larger for $i = 90^\circ$ and 40% larger for $i = 135^\circ$, compared to an aligned ($i = 0^\circ$) disk.

If the innermost gap-opening resonance lies outside of the tidal radius, then the disk is truncated non-resonantly. This tidal radius has been estimated to be $75 - 90\%$ of the Roche lobe radius for aligned disks (Paczyński 1977; Papaloizou & Pringle 1977; Paczyński & Rudak 1980; Pichardo, Sparke & Aguilar 2005). In the absence of an equivalent theory for misaligned disks, we estimate the

tidal radius to be $r_{L1}(e)$, the average Roche lobe radius at the binary pericenter separation [see equation (5.24)]. We find that a circumstellar disk in an equal mass binary fills its Roche lobe if inclined more than 45°, for our standard disk parameters ($h = 0.05$ and $\alpha = 0.01$). If αh^2 is smaller by a factor of $10^1 - 10^2$ (a reasonable range for protoplanetary disks), resonant truncation can restrict the disk size to less than $r_{L1}(e)$ for inclinations up to 90°. In non-equal mass binaries (we have considered $M_2/M_1 = 2.3 - 9.0$), for our standard disk parameters, the circumprimary disk fills its Roche lobe if inclined by more than 45°, while the circumsecondary disk barely fills its Roche lobe at an inclination of 90°. Since the Roche lobe formula [equation (5.24)] is approximate, 3D numerical calculations are needed to determine the precise outer radii of such disks.

We have also considered the resonance clearing of the inner cavity of a circumbinary disk. The most important resonances are the $\Omega/\Omega_B = 1{:}N$ ($N \geq 2$) and 2:3 commensurabilities. As for a circumstellar disk, the outer Lindblad torques have a complicated dependence on inclination and eccentricity. However, unlike for a circumstellar disk, the dependence of the inner disk truncation radius on eccentricity can be very different at large inclinations (see Figure 5.9). For an aligned disk, the inner cavity radius increases with eccentricity, and cannot extend past the 1:6 commensurability. This is also true at small inclinations (e.g., 22.5°), for which the cavity is slightly smaller than for an aligned disk. At larger inclinations, the cavity size can either be an increasing or decreasing function of eccentricity, possibly exhibiting both behaviors of different ranges of e. Nonetheless, the inner disk radius is generally smaller at large inclinations, for example, never extending past the 1:4 commensurability for $i = 45°$ or $i = 90°$, or past the 1:3 commensurability for $i = 135°$.

5.7.2 Approximations and Uncertainties

The calculations presented in this paper adopt several assumptions. We have assumed the disk to be flat. This is reasonable, as under typical conditions, bending waves and viscous stress lead to small disk warps (see Foucart & Lai 2014). Our results for the disk truncation radii (for both circumstellar and circumbinary disks) are based on the assumption of exactly Keplerian rotation profiles ($\Omega \propto r^{-3/2}$). The validity of this approximation is addressed in Section 5.6. We find that, although non-Keplerian effects due to the binary potential can modify the resonance properties significantly at high eccentricities (e.g., some resonances can be rendered non-existent), gap-clearing by the relevant resonances occurs at low eccentricities and is only slightly modified. Thus, non-Keplerian rotation effects have a negligible effect on disk truncation. We note, however, that we did not consider the effect of strong pressure gradients near the edge of the disk, which may cause further modification of the rotation profile and LRs (Petrovich & Rafikov 2012).

Our most important assumption (which is also made in AL94) is that angular momentum is deposited into the bulk disk material at, or very close to, the Lindblad resonances. This may affect our results in an appreciable way. First, angular momentum cannot be directly imparted to the disk material at Lindblad resonances, but must be carried as waves (Goldreich & Nicholson 1989), and received by the disk where the waves dissipate, either due to viscous damping (Takeuchi, Miyama & Lin 1996) or due to wave steepening and shock formation, although the latter should occur almost immediately after the waves are excited for the mass ratios we have considered (Goodman & Rafikov 2001). Second, the excited waves have broad angular momentum flux profiles, with

widths comparable to their radial wavelengths (e.g., Meyer-Vernet & Sicardy 1987), rather than being sharply peaked at the Lindblad resonances. For these reasons, the locations of disk truncation computed in this work may differ from those produced by self-consistent numerical treatments which account for these effects. For example, the hydrodynamic simulations of disks aligned with equal mass circular binaries by MacFadyen & Milosavljevic (2008) show that while the resonance responsible for clearing the inner cavity ($m = 2$, $N = 2$) is formally located at $r = 1.31a$ for a Keplerian disk, the actual cavity extends to nearly $2a$. This suggests that the details of wave excitation, propagation and breaking (or damping) are important uncertainties in the disk truncation process. With this is mind, our results for the truncation radii as functions of binary eccentricity and inclination should be interpreted as general trends rather than exact, sharply defined boundaries. We emphasize the need for detailed numerical work to fully interpret and assess the results of this paper.

CHAPTER 6

VISCOUS HYDRODYNAMICS SIMULATIONS OF CIRCUMBINARY ACCRETION DISKS: VARIABILITY, QUASI-STEADY STATE, AND ANGULAR MOMENTUM TRANSFER

6.1 Introduction

Circumbinary disks have been observed in a number of young stellar binaries (e.g., Dutrey et al. 1994; Mathieu et al. 1997; Simon et al. 2000). They are also expected to exist around supermassive black hole (SMBH) binaries as a consequence of accretion from the interstellar medium following galaxy mergers (e.g., Begelman et al. 1980; Ivanov et al. 1999; Milosavljević & Phinney 2005; Escala et al. 2005; Mayer et al. 2007; Dotti et al. 2007; Cuadra et al. 2009; Chapon et al. 2013). Understanding the dynamical behavior of circumbinary disks has important applications for a variety of astrophysical problems, including variable accretion in both young stellar objects (Jensen et al. 2007; Muzerolle et al. 2013; Bary & Petersen 2014) and active galactic nuclei (e.g., D'Orazio et al. 2015), circumbinary planet formation (Paardekooper et al. 2012; Meschiari 2012; Rafikov 2013; Silsbee & Rafikov 2015), and long-term binary orbital evolution (e.g., Armitage & Natarajan 2002; Haiman et al. 2009; Chang et al. 2010; Roedig et al. 2012). In this paper we carry out a suite of numerical simulations to investigate the structure, morphology, and variability of circumbinary accretion on a wide range of time-scales, as well as the long-term angular momentum transfer between the binary and the circumbinary disk.

This chapter is adapted from Miranda, Muñoz, & Lai (2017).

Several characteristic features distinguish circumbinary disks from circumsingle disks. A central low-density cavity, with a size of a few times larger than the binary orbit, is carved out around the binary by resonant gravitational torques, which prevent direct viscous inflow of mass (e.g., Artymowicz & Lubow 1994; Miranda & Lai 2015; note that binaries with small mass ratios open annular gaps, rather than common cavities, see D'Orazio et al. 2016). However, accretion of mass by the binary is by no means suppressed, and proceeds via narrow, non-axisymmetric streams which penetrate the central cavity (e.g., Artymowicz & Lubow 1996; MacFadyen & Milosavljević 2008; Muñoz & Lai 2016). These lead to time variability of the binary mass accretion rate. Streams of material may also be launched outward into the cavity by orbital motion of the binary (Muñoz & Lai 2016). Additionally, the binary excites eccentricity in the disk, possibly due to eccentric Lindblad resonances (Lubow 1991), and the eccentric disk may precess around the binary (e.g., Nelson 2003; Shi et al. 2012), potentially producing long-term variabilities (Muñoz & Lai 2016).

The variable mass accretion of the binary on orbital time-scales has been demonstrated both numerically (e.g., MacFadyen & Milosavljević 2008; Muñoz & Lai 2016) and observationally (Jensen et al. 2007; Muzerolle et al. 2013; Bary & Petersen 2014). Numerically, the accretion rate is found to vary primarily with a period equal to either the binary orbital period ($P_{\rm B}$), or about 5 times the binary orbital period. The occurrence of the longer period ($5P_{\rm B}$) variability is found to be linked to the properties of the binary. For example, it is suppressed for circular binaries with sufficiently small mass ratios (Farris et al. 2014). It is also suppressed when the binary is eccentric (Muñoz & Lai 2016). However, the detailed dependence of this behavior on the binary eccentricity has not been fully explored.

There have been conflicting claims in the literature about how the average mass accretion rate of the binary compares to that of a single star of the same mass. The ratio of the former to the latter characterizes the extent to which gravitational torques from the binary suppress or enhance accretion. Previous studies have reported values of this ratio both smaller than unity (e.g., MacFadyen & Milosavljević 2008; Ragusa et al. 2016), and greater than unity (e.g., Farris et al. 2014). In considering these claims, it is necessary to make a distinction between disks whose outer edges spread freely, and those whose outer regions are supplied with mass at a steady rate (e.g., due to an infalling envelope). For a freely spreading disk, a steady state, in which the accretion rate is constant throughout the disk, is never reached, and so the accretion rate of the binary continually changes with time (e.g., Rafikov 2016). For a disk with a steady mass supply at the outer region, a quasi-steady state is possible. In such a state, the time-averaged accretion rate of the binary must necessarily be equal to the rate at which mass is supplied to the outer disk, which is governed by processes occurring far from the binary. The time-averaged accretion rate of the central object must then be independent of whether it is a single body or a binary. The quasi-steady state was demonstrated in the simulations of Muñoz & Lai (2016) using the moving mesh code AREPO, which traces the gas from a circumbinary disk via accretion streams to circumsingle disks.

The rate at which angular momentum is transferred to the binary from the circumbinary disk is an outstanding open question. If there were no mass accretion onto the binary, the angular momentum coupling would be mediated only through gravitational torque, and the binary would lose angular momentum to the surrounding disk. It is generally thought that such angular momentum loss could play a central role in the mergers of SMBH binaries–shrinking

their orbits from $\sim$ 1pc to $\sim$ 0.01pc separations, at which gravitational radiation will cause them to merge within a Hubble time–providing a possible solution to the so-called "final parsec problem" (e.g., Begelman et al. 1980, Armitage & Natarajan 2002; Wyithe & Loeb 2003; Jaffe & Backer 2003; Sesana et al. 2008; MacFadyen & Milosavljević 2008; Haiman et al. 2009; Kelley et al. 2016). However, mass accretion carries positive angular momentum to the binary (in the case of a prograde disk). As a result of the competition between these two effects, the binary will, on average, either experience a net gain or suffer a net loss of angular momentum. The sign of this net angular momentum transfer has important consequences for the structure of the circumbinary disk, and directly determines (along with the rate of orbital energy transfer) the orbital evolution of the binary. As circumbinary accretion is generally variable on a range of timescales, controlled, long-duration numerical simulations are required to quantify the balance between gravitational torques and angular momentum advection.

In this paper, we present a series of 2D viscous hydrodynamics simulations of circumbinary accretion. While we adopt the simplest physical ingredients in our simulations (e.g., no magnetic fields or radiation transfer are included, and viscosity is prescribed using the α ansatz), our main goal is to carry out numerical experiments systematically in a well-controlled manner and for sufficiently long durations, so that reliable information can be obtained for the long-term evolution of the disk and for the net angular momentum transfer rate to the binary. To this end, we feed the outer disk boundary with a constant mass supply and ensure that the outer disk is close to the steady state. We carry out simulations over many viscous times of the "mid-disk" region (far away from the disk truncation radius), and ensure that a large region of the disk reaches a quasi-steady state. We survey a wide range of binary eccentricity, and to a

lesser extent, the binary mass ratio and disk viscosity parameter, and study how the disk behavior and evolution depends on these parameters. We find that, in addition to short-term variabilities of the binary accretion rate, the disk also exhibits long-term variabilities, corresponding to the coherent precession and apsidal locking of the inner disk. On average, the central binary can always accept mass at a rate very close to that supplied from the outer disk. Most importantly, we find that the net angular momentum received by the binary, including contributions from mass accretion, viscosity, and gravitational torque, depends on the binary eccentricity ($e_{\rm B}$), and is positive for a wide range of $e_{\rm B}$. This dependence of the angular momentum transfer on $e_{\rm B}$ correlates with the secular behaviors of the inner eccentric disk. Thus, contrary to the widely-held presumption that the binary loses angular momentum to the circumbinary disk, we find that when the effect of mass accretion is accounted for, the binary generally gains angular momentum to the extent that its semi-major axis tends to grow with time.

The plan for this paper is as follows. In Section 6.2, we describe our numerical setup. In Section 6.3, we examine the morphological features of the disk, and discuss the truncation of the inner cavity. Section 6.4 examines the variability of the mass accretion rate of the central binary on short (orbital) time-scales. In Section 6.5, we present numerical results on the long-term variation of the disk; such variation manifests as the precession and apsidal locking of the inner disk. In Section 6.6, which is self-contained, we explore possible theoretical explanations for the apsidal locking phenomenon. In Section 6.7, we investigate the long-term evolution of the global mass accretion rate and angular momentum accretion rate of the disk, and determine the average rate at which angular momentum is transferred to the binary. We summarize and discuss our results

in Section 6.8.

6.2 Problem Setup and Overview

A variety of computational approaches have been used to investigate the problem of circumbinary accretion, including both smoothed-particle hydrodynamics (e.g., Artymowicz & Lubow 1996; Cuadra et al. 2009; Roedig et al. 2012; Pelupessy & Portegies Zwart 2013; Dunhill et al. 2015) and Eulerian methods. There is a dichotomy in the Eulerian approaches between those which include the co-orbital region of the binary in the computational domain, and those which excise this region. When the co-orbital region is excluded (e.g., MacFadyen & Milosavljević 2008; Shi et al. 2012; D'Orazio et al. 2013; Shi & Krolik 2015; Lines et al. 2015), the flow can be solved efficiently using a polar grid geometry and an orbital advection algorithm (e.g., FARGO; Masset 2000) to subtract out the largely azimuthal average fluid motion, due to the fact that it is largely azimuthal. When the co-orbital region, where the flow is much less uniform, is included, the efficiency of this approach is lost. In this case, other methods, including the use of Cartesian grids (e.g., Günther & Kley 2002; Hanawa et al. 2010; de Val-Borro et al. 2011) or moving meshes (e.g., Farris et al. 2014; Muñoz & Lai 2016), have been adopted. In this paper, we employ the polar grid method with an excised binary co-orbital region. Although we do not follow the details of the flow around and onto the individual members of the binary, we take advantage of the lower computational cost in order to perform long-term (viscous time-scale) integrations, and to explore a wide variety of binary orbital parameters.

6.2.1 Numerical Setup

We solve the viscous hydrodynamic equations describing a thin disk around a binary, which consists of masses M_1 and M_2 (total mass $M_\mathrm{B} = M_1 + M_2$), with mass ratio $q_\mathrm{B} = M_2/M_1$, semi-major axis a_B, and eccentricity e_B. The positions (r_i, ϕ_i in polar coordinates) of M_1 and M_2 as a function of time are obtained by solving Kepler's equation using a higher-order extension of Newton's method (e.g., Murray & Dermott 1999). In our simulations, the circumbinary disk extends from $r_\mathrm{in} = (1 + e_\mathrm{B})a_\mathrm{B}$ to $r_\mathrm{out} = 70a_\mathrm{B}$, and is subject to the total gravitational potential of the binary,

$$\Phi(r, \phi, t) = -\sum_{i=1}^{2} \frac{GM_i}{\left[r^2 + r_i^2 - 2rr_i \cos(\phi - \phi_i)\right]^{1/2}}. \tag{6.1}$$

The equation of state is locally isothermal,

$$P = c_\mathrm{s}^2(r)\Sigma, \tag{6.2}$$

where $c_\mathrm{s}(r) = hr\Omega_\mathrm{K}$ is the sound speed and $\Omega_\mathrm{K} = (GM_\mathrm{B}/r^3)^{1/2}$. The disk aspect ratio, $h = H/r$ (where $H = c_\mathrm{s}/\Omega_\mathrm{K}$ is the pressure scale height) is therefore constant with r. Throughout this paper we choose $h = 0.1$. The kinematic viscosity is prescribed using the α-ansatz, $\nu = \alpha H c_\mathrm{s} = \alpha h^2 r^2 \Omega_\mathrm{K}$.

The initial surface density of the disk is

$$\Sigma(r) = \frac{\dot{M}_0}{3\pi\alpha h^2 \sqrt{GM_\mathrm{B} r}} \left[1 - \left(\frac{r_\mathrm{in}}{r}\right)^{1/2}\right] \exp\left[-\left(\frac{r}{r_\mathrm{edge}}\right)^{-2}\right], \tag{6.3}$$

and the initial radial velocity is

$$u_r(r) = -\frac{3\nu}{2r}\left[1 - \left(\frac{r_\mathrm{in}}{r}\right)^{1/2}\right]^{-1}. \tag{6.4}$$

For $r \gg r_\mathrm{edge}$, these profiles correspond to a steady state, constant $\dot{M}$ disk, with a zero torque condition at r_in. The Gaussian factor in Eq. (6.3) creates an artificial cavity around the central binary at the approximate radius of the real cavity

which is eventually self-consistently produced, in order to avoid violent relaxation at the beginning of the simulation. We set $r_{\rm edge} = 2r_{\rm in}$. The initial rotation profile of the disk is in centrifugal balance, including contributions from the binary quadrupole potential and pressure gradients,

$$\Omega^2(r) = \Omega_{\rm K}^2 \left[1 + \frac{3}{4} \frac{q_{\rm B}}{(1+q_{\rm B})^2} \left(1 + \frac{3}{2} e_{\rm B}^2 \right) \left(\frac{r}{a_{\rm B}} \right)^{-2} \right] + \frac{1}{r\Sigma} \frac{{\rm d}P}{{\rm d}r}. \tag{6.5}$$

At the outer boundary $r_{\rm out}$, all fluid variables are assumed to be fixed at their steady state values, and mass is injected into the domain at a rate $\dot{M}_0$. In code units ($GM_{\rm B} = a_{\rm B} = \Omega_{\rm B} = 1$), we choose $\dot{M}_0 = 3\pi\alpha h^2$, so that $\Sigma \approx r^{-1/2}(1 - \sqrt{r_{\rm in}/r})$ at $r \sim r_{\rm out}$. At the inner boundary, we employ a "diode" boundary condition, in which zero-gradient conditions ($\partial/\partial r = 0$) are imposed on Σ and u_ϕ, as well as on u_r, whenever it is negative. When u_r is positive, it is instead reflected across the boundary so that $u_r(r_{\rm in}) = 0$. This ensures that mass is allowed to leave the domain but cannot enter it through the inner boundary.

The fluid equations are solved using the finite volume, shock-capturing hydrodynamics code PLUTO (Mignone et al. 2007). We use third-order Runge-Kutta time stepping, piecewise parabolic spatial reconstruction, a Roe method Riemann solver, and the FARGO orbital advection algorithm. We use a polar grid centred on the centre of mass of the binary, with uniform grid spacing in the azimuthal direction and logarithmic grid spacing in the radial direction. Unless otherwise stated, $N_\phi = 600$ azimuthal grid cells are used. The number of radial cells is chosen so that $\Delta r \approx r\Delta\phi$, i.e., the cells are approximately square, by setting $N_r \approx [N_\phi/(2\pi)] \ln(r_{\rm out}/r_{\rm in})$. All runs have the same value of $r_{\rm out} (= 70a_{\rm B})$, but different values of $r_{\rm in} = (1 + e_{\rm B})a_{\rm B}$, and therefore different values of N_r.

q_B	e_B	α	t_{end} [P_B]	$r_{rel}(t_{end})$ [a_B]	r_{peak} [a_B]	$r_{0.1}$ [a_B]	Lump	$\dot{\varpi}_d$ [Ω_B]	l_0 [$a_B^2\Omega_B$]
1.0	0.0	0.1	3000	12.16	3.56	1.67	Yes	3.38×10^{-3}	$+0.816 \pm 0.027$
1.0	0.05	0.1	3000	12.16	3.52	1.66	Yes	3.32×10^{-3}	$+0.825 \pm 0.025$
1.0	0.1	0.1	3000	12.16	3.64	1.65	No	7.64×10^{-4}	$+0.690 \pm 0.098$
1.0	0.2	0.1	5000	17.10	3.40	1.80	No	None	$+0.674 \pm 0.021$
1.0	0.4	0.1	8000	23.39	3.15	2.02	No	None	$+0.273 \pm 0.017$
1.0	0.6	0.1	5000	17.10	4.58	2.23	No	1.92×10^{-3}	$+1.053 \pm 0.018$
1.0	0.8	0.1	4000	14.73	5.27	2.34	No	1.79×10^{-3}	$+1.249 \pm 0.012$
1.0	0.0	0.05	7000	13.48	3.95	1.88	Yes	2.71×10^{-3}	$+0.800 \pm 0.061$
1.0	0.2	0.05	25000	31.49	3.13	1.95	No	None	-0.481 ± 0.145
1.0	0.3	0.05	10000	17.10	4.63	2.18	No	1.44×10^{-3}	$+0.996 \pm 0.056$
1.0	0.4	0.05	12000	19.31	4.49	2.28	No	1.80×10^{-3}	$+0.971 \pm 0.053$
1.0	0.8	0.05	10000	17.10	5.33	2.57	No	1.60×10^{-3}	$+1.196 \pm 0.032$
0.5	0.0	0.1	3000	12.16	3.45	1.63	Yes	3.18×10^{-3}	$+0.830 \pm 0.043$
0.5	0.4	0.1	5000	17.10	4.35	1.96	No	1.44×10^{-3}	$+1.030 \pm 0.028$
0.2	0.0	0.1	3000	12.16	3.08	1.42	No	3.33×10^{-3}	$+0.873 \pm 0.020$
0.2	0.4	0.1	3000	12.16	4.68	1.69	No	None	$+1.119 \pm 0.054$

Table 6.1: Summary of parameters and key results of all simulations presented in this paper. The first five columns give the binary mass ratio q_B and eccentricity e_B, the disk viscosity parameter α, the total integration time t_{end}, and the viscous relaxation radius at t_{end}. The next two columns are the two radii characterizing the inner disk truncation, r_{peak} and $r_{0.1}$. The column labelled "Lump" indicates the presence or absence of a lump feature which causes variability of the mass accretion rate with a period of $\sim 5P_B$. The second to last column gives the precession frequency of the eccentric inner disk, or indicates if it is instead apsidally aligned with the binary (if "None" is listed). The last column is the net angular momentum received by the binary per unit accreted mass.

6.2.2 Analysis Procedure

Table 6.1 summarizes the parameters and key results of all the simulations presented in this paper. The relevant quantities and results are discussed in the main sections of the paper. Of primary interest to our study is allowing the disk to reach a steady state. This requires that the influence of the central binary be communicated, by viscosity, to a sufficiently large radius. The viscous timescale at r is $t_\nu = (4/9)(r^2/\nu)$ (Lynden-Bell & Pringle 1974). Therefore, after a time t, the disk is viscously relaxed within a radius r_{rel}, defined by $t = t_\nu(r_{rel})$:

$$r_{rel}(t) = \left(\frac{9\pi}{2}\alpha h^2 \frac{t}{P_B}\right)^{2/3} a_B. \tag{6.6}$$

Table 6.1 indicates the total integration time $t_{\rm end}$ of each simulation, and its corresponding viscous relaxation radius $r_{\rm rel}(t_{\rm end})$. In all cases, we allow the disk to relax out to at least $12a_{\rm B}$, which is significantly larger than the inner disk radius. However, this is only a necessary, not a sufficient condition, for the disk to be in a quasi-steady state. In Section 6.7, we discuss additional criteria we use to determine that the disk has fully relaxed.

During the long-term evolution ($t = [0, t_{\rm end}]$) of each simulation, the fluid variables are output every $10P_{\rm B}$. This information is used to determine the secular evolution of the orbital elements of the disk, in particular, whether the disk precesses or is apsidally aligned with the binary (see Section 6.5). Several shorter duration runs are restarted from snapshots taken during the long-term evolution. The duration of the shorter runs are chosen according to the long-term behavior of the orbital elements of the disk. For precessing disks, the duration is equal to the precession period (as this is the longest period over which variability occurs), to the nearest $50P_{\rm B}$, and the fluid variables are output 5 times per $P_{\rm B}$. For the aligned cases, the restarted runs last only $50P_{\rm B}$, and the variables are output 20 times per $P_{\rm B}$. In either case, at least 1000, and as many as 3500 snapshots of the fluid variables are captured in each restart run. The snapshots are used to compute time-averaged profiles of the mass accretion rate $\dot{M}(r)$, and the net angular momentum accretion rate $\dot{J}(r)$ (see Section 6.7 and Eq. 6.28). We use the evolution of these quantities to verify that the simulations are evolving towards a quasi-steady state, as described in Section 6.7. Finally, each simulation is restarted from $t_{\rm end}$, and evolved for an additional $50P_{\rm B}$, with a sampling rate of $20/P_{\rm B}$. This restart run is used to compute the time-averaged surface density profiles shown in Section 6.3, as well as the time series of $\dot{M}$ shown in Section 6.4. These analyses are therefore performed while the disk has reached

the quasi-steady state.

In order to check the effects of numerical resolution, we also performed two higher resolution runs, with N_r and N_ϕ both twice as large as in the standard runs, for the cases $(q_{\rm B}, e_{\rm B}, \alpha) = (1.0, 0.4, 0.1)$ and $(1.0, 0.2, 0.05)$. Due to their high computational cost, these were only evolved for $1000P_{\rm B}$, followed by the $50P_{\rm B}$ restart runs. We verified that the features and behaviors found in these runs were largely unchanged compared to their standard resolution counterparts.

6.3 Disk Morphology

6.3.1 General Features

Figure 6.1 shows a sample of surface density snapshots for each of the simulation runs (see Table 6.1). These snapshots are taken near the end of each simulation, when the inner disk has reached a quasi-steady state. Only the innermost part of the disk, where there are strong visible deviations from axisymmetry, is shown. The snapshots illustrate several features that are common to all of the simuations. There is a low-density central cavity, approximately $(2-3)a_{\rm B}$ in radius, carved out by gravitational torques exerted by the binary (see Section 6.3.2). Spiral density waves, primarily with azimuthal number $m = 1$ or 2, are excited by the binary and propagate outwards to several times the cavity radius. The central cavity is asymmetric, and is penetrated by dense (relative to the low average surface density in the cavity), narrow accretion streams which carry mass toward the inner boundary. The disk is eccentric, as seen by the shape of the cavity edge (see Section 6.5 for a detailed discussion of the eccentricity

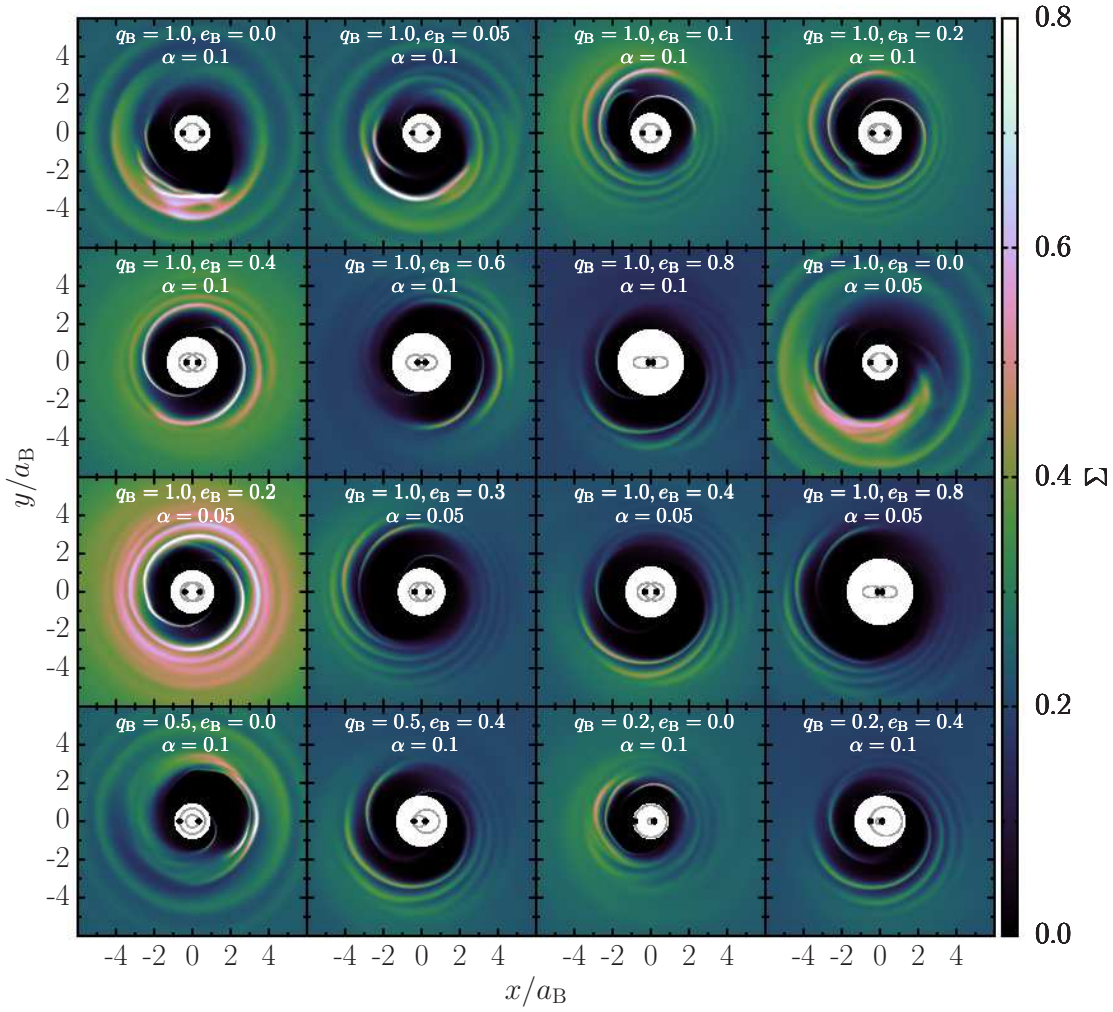

Figure 6.1: Snapshots of surface density in the inner disk after reaching a quasi-steady state (see Section 6.3.1). The snapshot for each simulation is taken at its corresponding time $t_{\rm end}$, as given in Table 6.1, at which the binary is always at pericentre. The binary orbit is shown in the centre of each panel. Note that for $e_{\rm B} = 0$ (and, to a lesser extent, $e_{\rm B} = 0.05$), an $m = 1$ lump develops near the disk inner edge.

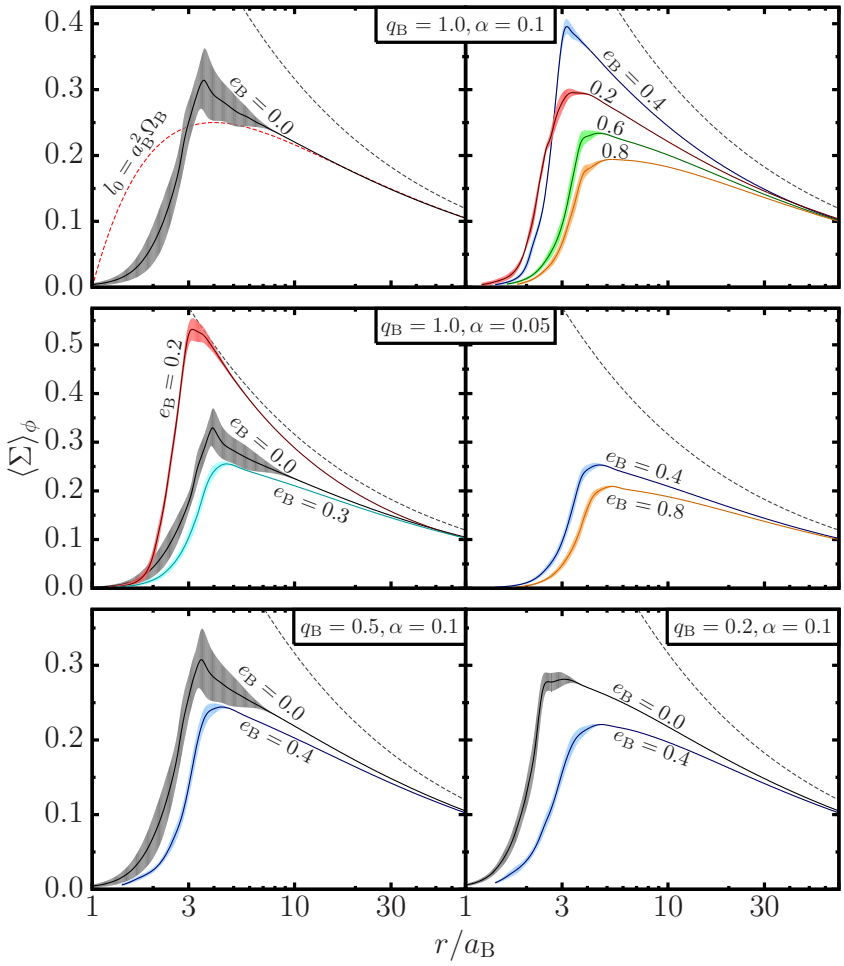

Figure 6.2: Azimuthally-averaged surface density profiles for various binary and disk parameters (see Section 6.3.2). The solid lines are time-averaged over 50 binary orbits, and the shaded regions around them indicate their 1σ variations. The black dashed line depicts the asymptotic surface density profile for $\sqrt{r/a_B} \gg 1$, $\Sigma = \dot{M}_0/(3\pi\nu)$. The red dashed line in the top-left panel shows the deviation from this profile in the outer disk due to the imposed value of l_0 (see Eq. 6.3) for $e_B = 0$.

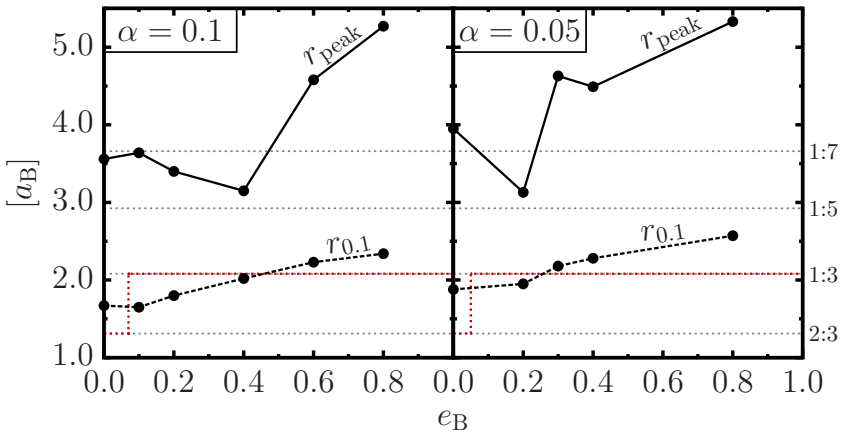

Figure 6.3: The radial locations of two points characterizing the truncation of the inner disk (see Section 6.3.2): $r_{\rm peak}$, the point at which the azimuthally averaged surface density is largest, and $r_{0.1}$, where it drops to 10 per cent of the value at $r_{\rm peak}$, as a function of binary eccentricity. Only equal-mass binaries ($q_{\rm B} = 1.0$) are shown. The locations of several resonances which are important to disk truncation are shown on the right, and the dashed red line shows the theoretical predictions for the truncation radius as derived by Miranda & Lai (2015).

dynamics of the inner disk). Finally, for circular, or nearly circular ($e_{\rm B} < 0.1$) binaries only, there is a overdense lump, i.e., radially localized feature with approximate $m = 1$ symmetry, at the edge of the cavity. The lump orbits the binary with approximately the local Keplerian orbital period, while exhibiting a continuous cycle of creation and destruction with the same period. This process is associated with modulation of the mass accretion rate through the inner boundary at the frequency $\sim \Omega_{\rm B}/5$ (see Section 6.4).

6.3.2 Inner Disk Truncation

Figure 6.2 shows the double-averaged (i.e., azimuthally-averaged and time-averaged) surface density profiles for most of the runs depicted in Fig. 6.1. Generically, $\langle\Sigma\rangle$ has a positive slope ($\mathrm{d}\langle\Sigma\rangle/\mathrm{d}r > 0$) in the inner disk (within a

few a_B's from the binary), then reaches a maximum at a few a_B's and has negative slope beyond that. For some parameters, the maximum manifests as a sharp peak, while for others, it is broader and less well-defined. This is associated with variation in the value of the specific angular momentum eigenvalue l_0 for each simulation (see Section 6.7), with smaller values of l_0 corresponding to sharper peaks. The asymptotic behavior of $\langle\Sigma\rangle$ for $\sqrt{r/a_B} \gg 1$ is $\langle\Sigma\rangle = \dot{M}/(3\pi\nu) \propto r^{-1/2}$, which all of the profiles become similar to at large r, is also shown in Fig. 6.2. However, the density profiles do not exactly approach this form, due to the initial "guess" for l_0 imposed by the surface density profile (Eq. 6.3), $l_0^{\rm guess} = \sqrt{GM_B r_{\rm in}}$. This is demonstrated in the top left panel, for the case $(q_B, e_B, \alpha) = (1.0, 0.0, 0.1)$. Here the initial profile for the outer disk, characterized by $l_0 = 1$ (in units of $a_B^2\Omega_B$) is shown. The actual surface density profile remains nearly identical to this for $r \gtrsim 10a_B$, since the influence of the binary has not yet been viscously communicated to this region.

Regardless of its sharpness, $\langle\Sigma\rangle$ always has a maximum, typically located at $r/a_B \sim 3-5$. We denote this radius $r_{\rm peak}$. Between $r_{\rm peak}$ and the inner boundary of our simulation domain [$r_{\rm in} = (1 + e_B)a_B$], $\langle\Sigma\rangle$ drops by about two orders of magnitude, i.e., the disk is truncated interior to $r_{\rm peak}$. Specifying the exact location of the "truncation radius" is ambiguous, but we define it as the radius (denoted by $r_{0.1}$) at which $\langle\Sigma\rangle$ is 10 per cent of its value at $r_{\rm peak}$.

Figure 6.3 shows $r_{\rm peak}$ and $r_{0.1}$, as a function of binary eccentricity. While $r_{\rm peak}$ is not monotonic in e_B (this is also associated with variations in the value of l_0), $r_{0.1}$ strictly increases with e_B. In the theory of inner disk truncation (Artymowicz & Lubow 1994; Miranda & Lai 2015), the truncation radius increases with e_B, as the torques applied at increasingly higher order resonances overcome the

viscous torque at those locations. The locations of the relevant resonances are shown in Fig. 6.3, as well the theoretical prediction for the truncation radius, using the formalism of Miranda & Lai (2015). The theoretical prediction is very similar for both values of α shown. In both cases, the $\Omega/\Omega_B = 2/3$ resonance is responsible for disk truncation for $e_B = 0$, and then, above a small critical eccentricity (which is slightly different for the two values of α), the 1/3 resonance becomes sufficiently strong to truncate the disk. Higher order resonances are never strong enough to truncate the disk for the viscosity parameters ($\alpha = 0.05$ and 0.1) considered in this paper. This discrete behavior is not observed in our simulations, given the "fuzziness" of the disk truncation. The theory predicts that the truncation radius lies between 1.3 and $2.1a_B$, while we find that $r_{0.1}$ is between 1.7 and $2.6a_B$. Thus, there is agreement at the 20 per cent level between the theory and our numerical results.

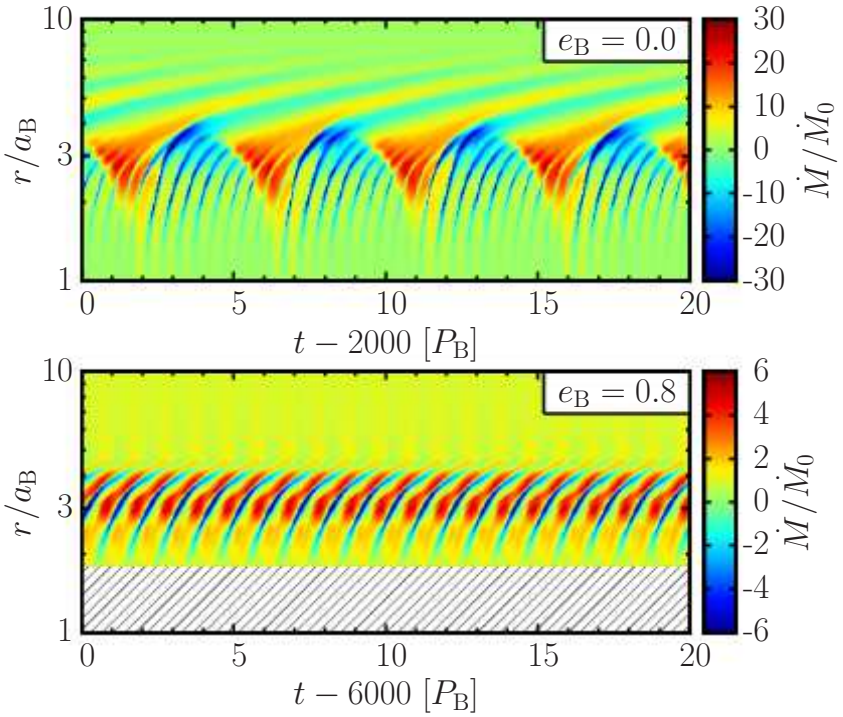

Figure 6.4: The mass accretion rate $\dot{M}(r,t)$ as a function of time (x-axis) and radius (y-axis), for $q_{\rm B} = 1.0$, $\alpha = 0.1$, and two different binary eccentricities, $e_{\rm B} = 0.0$ (top) and 0.8 (bottom). The accretion rate (colours) is normalized by the supplied rate at $r_{\rm out}$, $\dot{M}_0$. Note that the colour bars have different scales in the top and bottom panels. The hatched region in the bottom panel indicates that the computational domain does not extend down to $r/a_{\rm B} = 1$ as it does in the top panel. For $e_{\rm B} = 0.0$, the main periodicities that can be seen are at $5P_{\rm B}$ and $P_{\rm B}/2$. For $e_{\rm B} = 0.8$, the main variability is at $P_{\rm B}$. We find that this behavior is typical for eccentric binaries. In both cases, the accretion rate is steady beyond about $10a_{\rm B}$. This outer region is not shown, in order to focus on the variability of the inner disk (see Section 6.4.1).

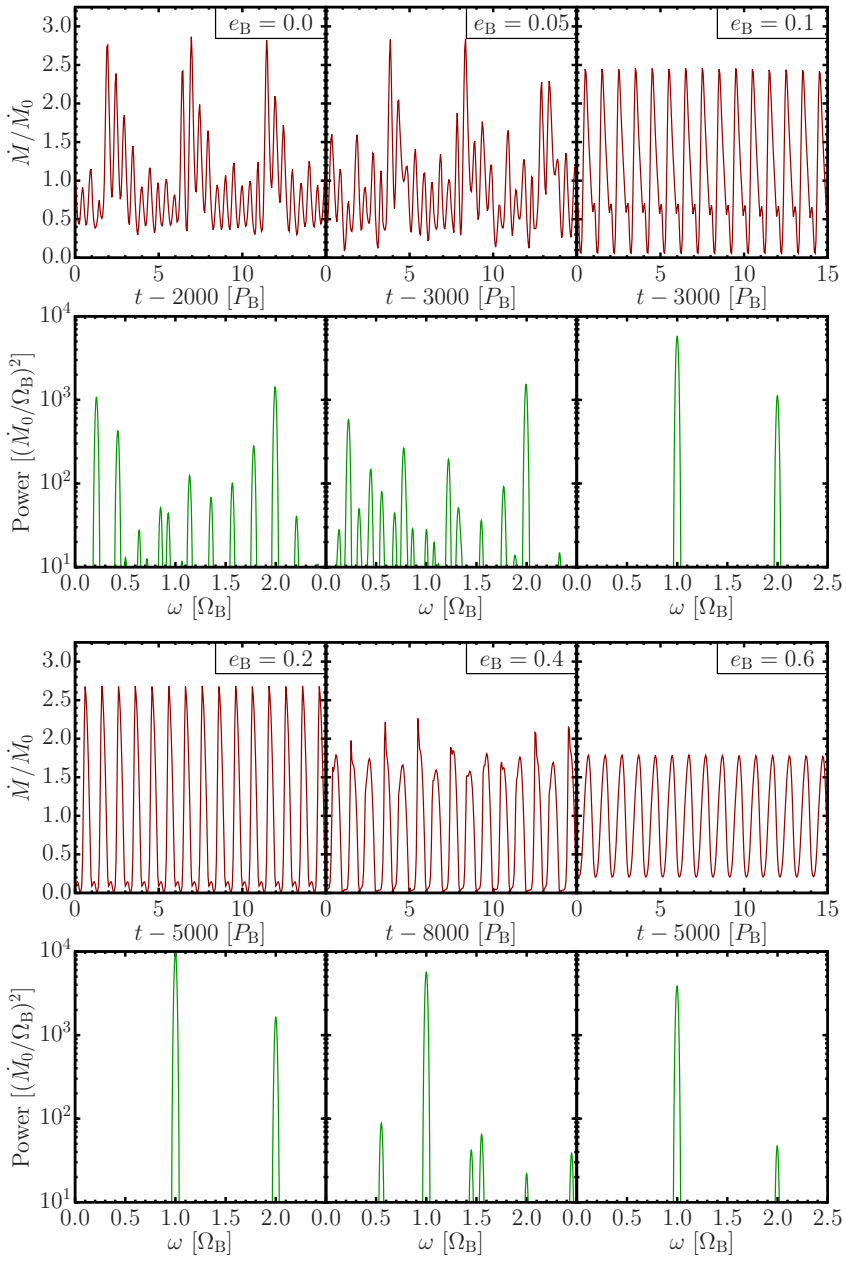

Figure 6.5: The mass accretion rate at $r_{in} = (1 + e_B)a_B$ for an equal mass binary ($q_B = 1.0$) and $\alpha = 0.1$, for a range of binary eccentricities, e_B (first and third rows), and its power spectrum, taken over $50P_B$ (second and fourth rows). The top sequence of eccentricities demonstrates the transition from accretion modulated at a frequency of approximately $\Omega_B/5$, to accretion modulated at Ω_B, which occurs at about $e_B = 0.05$. The bottom sequence indicates the typical behavior for the moderate to high binary eccentricity cases, where the accretion is modulated only at Ω_B. See Section 6.4.1.

6.4 Short Time-Scale Variability

6.4.1 Dependence on Binary Eccentricity

We explore the time-dependence of the local mass accretion rate[1],

$$\dot{M}(r,t) = -\oint r\Sigma u_r \mathrm{d}\phi, \tag{6.7}$$

which is generally variable on orbital time-scales. The accretion rate as a function of r and t are shown in Fig. 6.4, for two representative cases ($e_{\rm B} = 0$ and 0.8). For the circular case, two clear periodicities are evident. First, there are variations with a frequency of $2\Omega_{\rm B}$ in the inner disk ($r \lesssim 4a_{\rm B}$). Second, there are larger variations, with a frequency of about $\Omega_{\rm B}/5$, which are present out to about $\sim 10a_{\rm B}$, having the largest amplitude at about $(2-3)a_{\rm B}$. The second type of variability has been seen in previous simulations of disks around circular binaries (e.g., MacFadyen & Mirosavljevic 2008; Farris et al. 2014), and is sometimes referred to as having a frequency of $(2/9)\Omega_{\rm B}$ (note that we find the exact frequency to be $0.21\Omega_{\rm B}$, with a FWHM in power of $0.03\Omega_{\rm B}$, so the distinction between 1/5 and 2/9 is irrelevant). Note that in our simulations, $\dot{M}$ is always positive at the inner boundary, as required by the diode boundary condition.

[1] In this section, we use an approximation for $\dot{M}$, in which the (primitive) fluid variables $(\Sigma, \mathbf{u})$ are taken to be equal to their cell-averaged values at cell centres. These values, as output by PLUTO, are derived from the cell-averaged values of the conservative variables (Σ and momentum density $\mathbf{m}$). In other words, the inter-cell variation of the primitive variables, which vanishes in the limit of infinite resolution, is ignored. This approximation is valid for exploring the short-term time-dependence of $\dot{M}$, because its variations are much larger than its time-averaged value. However, when evaluating small fluctuations in the time-averaged profile of $\dot{M}$ (as well as that of $\dot{J}$), as in Section 6.7, the approximation breaks down. In this case, it is necessary to reconstruct the inter-cell variation, using the piecewise parabolic interpolation method (Colella & Woodward 1984), as employed in the hydrodynamic solver, to evaluate the fluid variables at cell interfaces, before taking their appropriate products to calculate $\dot{M}$ or $\dot{J}$. Taking products of the cell-averaged fluid variables can result in anomalous features in the profiles of $\dot{M}$ and $\dot{J}$, wherever the variables have strong inter-cell gradients. The interpolation allows the different variables to be evaluated at the same location before their products are taken, resulting in more accurate profiles.

Muñoz & Lai (2016) showed that this feature is preserved in more realistic simulations that resolve the accretion streams onto individual bodies, at least for the $e_B = 0$ case. Beyond about $10a_B$, $\dot{M} \approx \dot{M}_0$ at all times. From just beyond the inner boundary to about $10a_B$, the sign of $\dot{M}$ continuously alternates with tiime, indicating that, instantaneously, mass may be flowing inwards (towards the binary) or outwards (away from it). In a time-averaged sense, $\dot{M}$ is positive everywhere, so mass flows toward the binary on average (see Section 6.7). For the $e_B = 0.8$ case (see the bottom panel of Fig. 6.4), $\dot{M}$ varies only at Ω_B, and its maximum amplitude is about 5 times smaller than for the circular case. As in the case of the circular binary, $\dot{M}$ is strictly positive near the inner boundary, as well as sufficiently far from the binary ($r \gtrsim 6a_B$), but can have either positive or negative sign at intermediate distances. This "sloshing" effect (fluctuation in the magnitude and sign of $\dot{M}$), which is largest at $\sim 3a_B$, and largest overall for circular binaries, is responsible for the variations in the azimuthally-averaged surface density profiles shown in Fig. 6.2.

Of particular interest is the variability of the mass accretion rate at r_{in}, which we take to represent the actual accretion rate onto the binary (this is subject to the validity of the diode boundary condition). The different behaviors are illustrated in Fig. 6.5, which shows $\dot{M}(r_{in}, t)$, and its power spectrum, for equal mass binaries with various eccentricities. The top row demonstrates the transition from the circular binary behavior to the eccentric binary behavior. For $e_B = 0$, the same variabilities seen in Fig. 6.4 are evident: large amplitude "bursty" fluctuations at about $\Omega_B/5$, and smaller amplitude fluctuations at $2\Omega_B$. For $e_B = 0.05$, the behavior is similar to that of the $e_B = 0$ case, although the $\Omega_B/5$ spike in the power spectrum is somewhat less distinct, and additional low-frequency components are present. For $e_B = 0.1$, the variability behavior becomes drastically

different, and is dominated by oscillations at Ω_B and $2\Omega_B$. The power spectrum is much cleaner than for the lower values of e_B, only having peaks at these two frequencies. Therefore, the accretion rate only exhibits low-frequency (sub-Ω_B) variability for $e_B \lesssim 0.05$. The bottom two panels of Fig. 6.5 demonstrate that the simple variability at only Ω_B is the typical behavior for all higher values of e_B. The exception is the case of $e_B = 0.4$, for which the accretion rate variability appears less clean, and has small additional components present in the power spectrum (at about 0.5 and 1.5 times Ω_B). This intermediate e_B is also associated with other unique behaviors, including periapse locking of the disk with the binary (see Section 6.5), large gravitational torques, and low level of angular momentum transfer to the binary (see Section 6.7).

6.4.2 Lump at Inner Edge of Disk

The variability of $\dot{M}(r_{in}, t)$ for circular and slightly eccentric binaries ($e_B \lesssim 0.05$) is associated with the presence of a lump at the inner edge of the disk, which is not present for higher values of e_B. The periodic creation and destruction of this lump, with a period corresponding to the Keplerian orbital period at about $2.8a_B$, is temporally coincident with the modulation of $\dot{M}(r_{in}, t)$ with the same period (MacFadyen & Milosavljević 2008; Shi et al. 2012; D'Orazio et al. 2013; Farris et al. 2014; Muñoz & Lai 2016). The lump cycle is illustrated in the left two panels of Fig. 6.6 (note that the time interval shown here corresponds to the first 5 orbits shown in the leftmost panel of Fig. 6.5). After its creation, the lump rotates for about a third of an orbit before it is torn apart and flung towards the inner boundary. A new lump is then created by the accumulation of streams from the central cavity, and the cycle repeats. In Fig. 6.1, a lump,

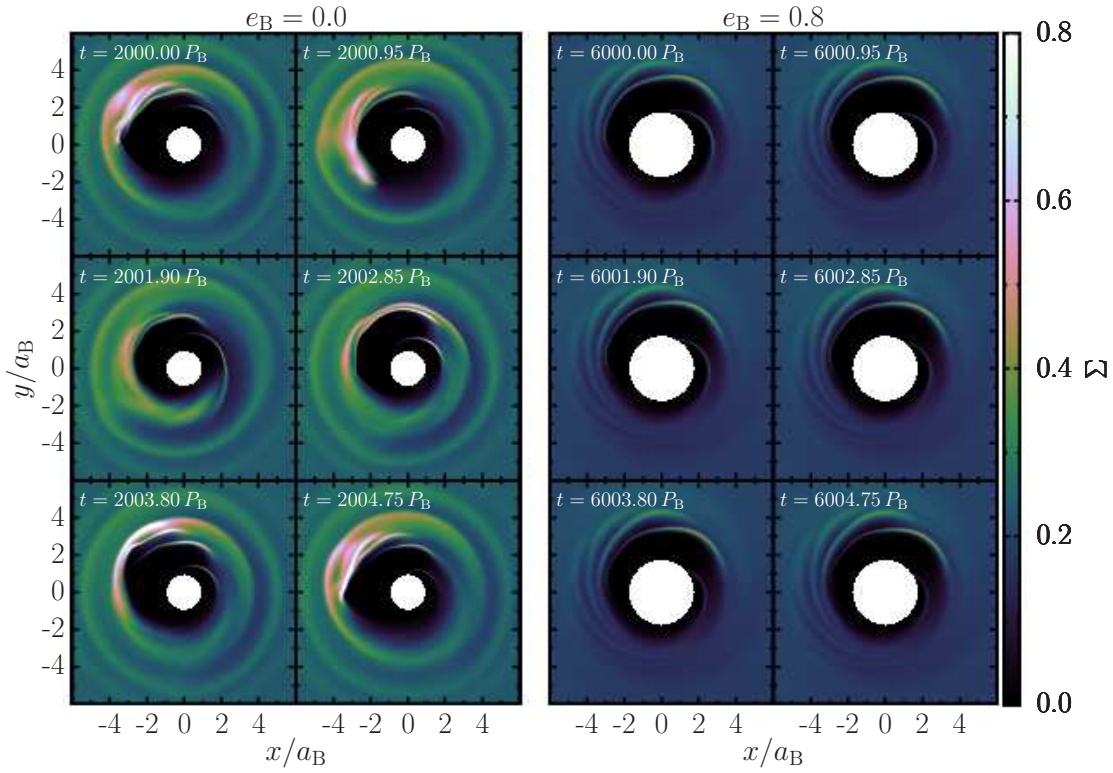

Figure 6.6: *Left*: Surface density snapshots illustrating one cycle of the creation and destruction of the lump at the inner edge of the disk for a circular binary. The six snapshots are evenly separated by 1/5 of the approximate period of the lowest frequency component of the power spectrum of $\dot{M}(r_{\rm in})$ (the exact period is $4.72P_{\rm B}$). In the first two panels, the lump is visible, orbiting at approximately the local Keplerian frequency. In the next panel, it has been sheared apart, and only a trace of it is left. This panel corresponds to the time at which the mass accretion rate onto the binary reaches its peak. In the fourth and fifth panels, the lump can be seen re-forming due to a pileup of streams emanating from the central cavity. Finally, in the last panel, the lump has reappeared, and is very close to its original position shown in the first panel. *Right*: Same as the left panels except for an $e_{\rm B} = 0.8$ binary. The relative steadiness of the disk over the same interval of time can be clearly seen. Spiral density waves, which pile up at the disk apocentre (near the top of each panel), are also visible. They appear very similar in different snapshots, which are separated by approximately one binary orbital period. However, small differences of the phase of the innermost density wave/accretion stream can be seen because they are not exactly one orbit apart. No sign of a lump can be seen in this case. The presence or absence of a lump correlates with the presence or absence of $\Omega_{\rm B}/5$ modulation of the accretion rate. See Section 6.4.2.

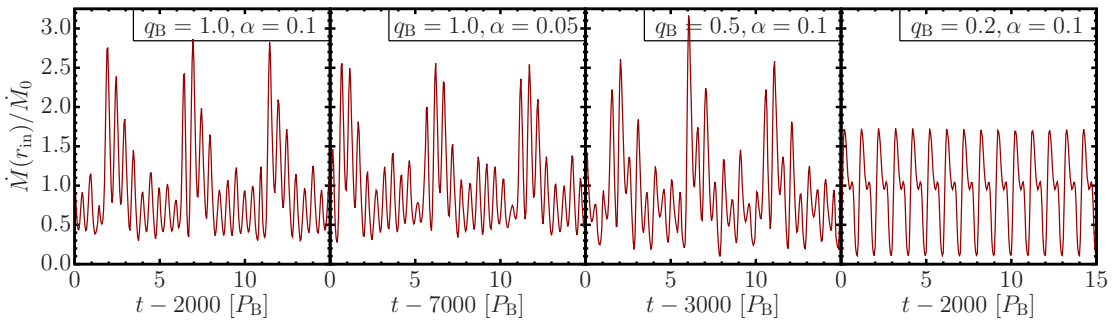

Figure 6.7: Mass accretion rate as a function of time (as in Fig. 6.5, but without the associated power spectra) for circular binaries ($e_B = 0$), for different mass ratios, and different values of α. The leftmost panel, with $q_B = 1.0$ and $\alpha = 0.1$, is repeated from Fig. 6.5. For equal mass binaries, low-frequency ($\Omega_B/5$) accretion modulation occurs indepedently of the value of α. For non-equal mass binaries, it also occurs, although somewhat modified, for $q_B = 0.5$, but is gone for $q_B = 0.2$, and is replaced instead by modulation at Ω_B only. See Section 6.4.3.

though slightly less well-defined, can also be seen for $e_B = 0.05$, which also displays low-frequency ($\sim \Omega_B/5$) accretion variability. For comparison, the right two panels of Fig. 6.6 show the disk around an eccentric binary over the same interval of time as shown in the left two panels. The lump feature is clearly not present. Only spiral density waves and accretion streams, which look remarkably similar from one orbit to the next, can be seen. This demonstrates why the variability of the accretion rate at r_{in} for eccentric binaries is essentially a clean sinusoid whose period matches that of the binary.

The fact that a lump appears in simulations with $e_B \lesssim 0.05$, but not for larger values of e_B, indicates that it is a sensitive structure. Disks around circular binaries experience gravitational forcings which have pattern frequencies equal to the binary orbital frequency only. For eccentric binaries, there are many additional forcings, having both larger and smaller pattern frequencies, which become stronger with increasing e_B. The fact that the lump disappears above a very small value of e_B may indicate that it is easily destroyed (or prevented

from forming) by these additional forcings.

6.4.3 Dependence on Binary Mass Ratio

Figure 6.7 shows $\dot{M}(r_{\rm in}, t)$ for simulations with circular binaries, with different values of $q_{\rm B}$ and α. They all exhibit low-frequency ($\sim \Omega_{\rm B}/5$) variabilities, except for the case of $q_{\rm B} = 0.2$, whose accretion rate varies only at a frequency of $\Omega_{\rm B}$, with weak modulation at $2\Omega_{\rm B}$, somewhat resembling the behavior of an eccentric binary. Figure 6.1 shows that this is also the only disk around a circular binary in our runs that does not have a lump feature. Therefore, the presence of a lump and associated low-frequency accretion variability not only requires a small binary eccentricity, but also a sufficiently large mass ratio ($q_{\rm B} \gtrsim 0.2$). This dependence on mass ratio was also pointed out by Farris et al. (2014).

6.5 Long Time-Scale Variability: Disk Eccentricity and Precession

Having studied the short-term ($\sim P_{\rm B}$) variabilities of the circumbinary accretion in Section 6.4, we now investigate variabilities of the disk on time-scales $\gg P_{\rm B}$. To this end, we examine the $m = 1$ Fourier component of the radial velocity u_r, and compute its power spectrum,

$$|\tilde{u}_{r,m=1}|^2(r,\omega) = \left| \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \oint u_r(r,\phi,t) \mathrm{e}^{\mathrm{i}(\phi-\omega t)} \mathrm{d}\phi \mathrm{d}t \right|^2 , \tag{6.8}$$

where $t_1 = t_{\rm end} - 2000 P_{\rm B}$ to $t_2 = t_{\rm end}$. Figure 6.8 shows the results for three different values of $e_{\rm B}$. We choose $u_{r,m=1}$ as a diagnosis of an eccentric disk, because an

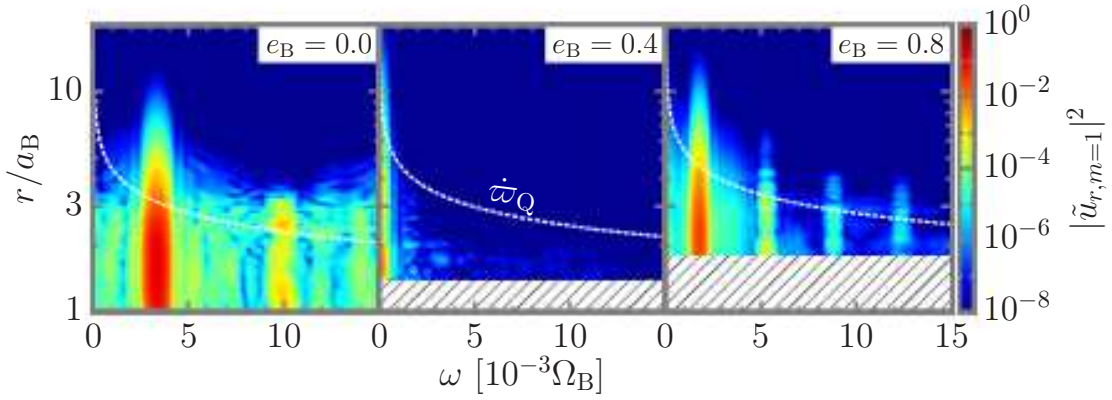

Figure 6.8: Low-frequency power spectrum of the $m = 1$ Fourier component of u_r, taken over 2000 binary orbits, for three different binary eccentricities (with $q_{\rm B} = 1.0$ and $\alpha = 0.1$ in all three cases). The dashed line in each panel corresponds to the precession frequency of a test particle due to the binary quadrupole potential, $\dot{\varpi}_{\rm Q}$ (Eq. 6.9). The presence of spatially coherent power for $e_{\rm B} = 0.0$ and $e_{\rm B} = 0.8$ corresponds to a global precessional mode. Such a mode is conspicuously absent for $e_{\rm B} = 0.4$ (note that the apparent power near zero frequency is spurious, as it corresponds to periods longer than the interval over which the power spectrum was taken). In the cases for which this mode is present, its frequency corresponds to $\dot{\varpi}_{\rm Q}$ at $r \approx (3-5)a_{\rm B}$, the approximate location of $r_{\rm peak}$.

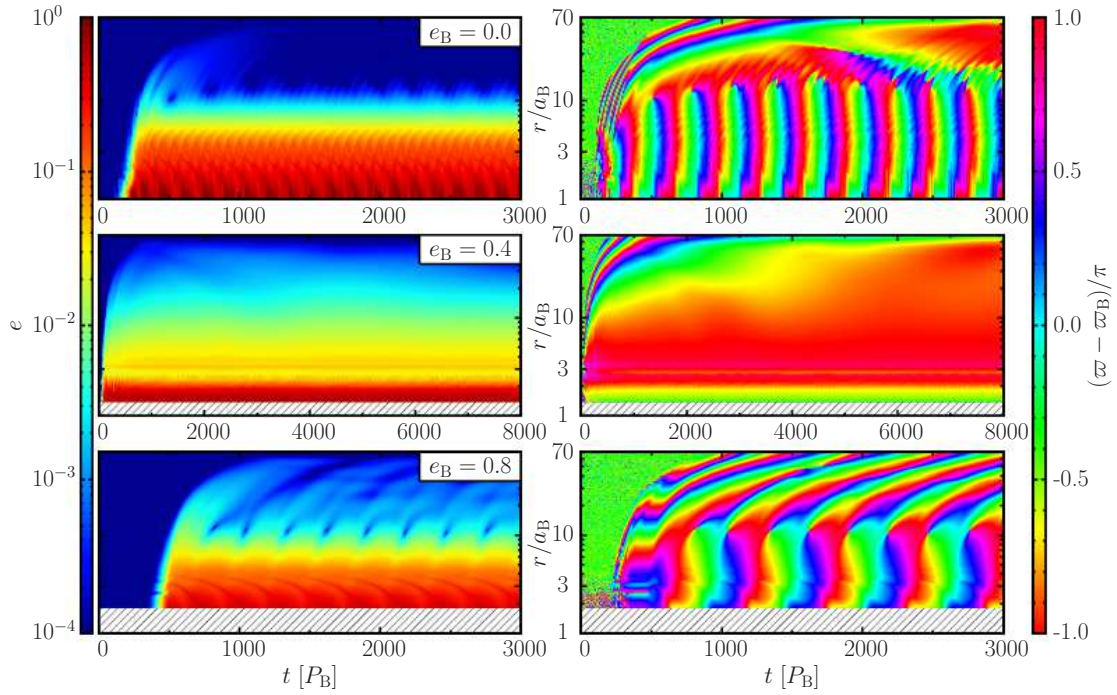

Figure 6.9: Examples of the two different behaviors of the disk eccentricity (precession and apsidal locking) for different binary eccentricities (see Section 6.5). The disk eccentricity (magnitude of the mass weighted, azimuthally averaged eccentricity vector; left panels) and argument of pericentre (phase of the mass weighted, azimuthally averaged eccentricity vector; right panels) are shown as a function of time (x-axis) and radius (y-axis), for different values of e_B (0, 0.4, and 0.8; different rows), all for $q_B = 1$ and $\alpha = 0.1$. The hatching in the bottom of the plots in the middle and bottom rows indicate that the computational domain does not extend all the way to $r/a_B = 1$ (instead, the inner disk edge is at $r_{in}/a_B = 1 + e_B$). Coherent precession of the inner eccentric disk [$(2 - 10)a_B$], as indicated by vertical strips in the right panels, is clearly seen for both $e_B = 0$ and 0.8. For $e_B = 0.4$, the pericentre of the inner eccentric disk instead remains aligned with the binary (note that $\varpi - \varpi_B$ equal to either 0 or $\pm\pi$ indicates alignment, since the two values are degenerate for the equal mass binary shown here). Note that the range of the time axis is the same in the top and bottom panels (thus the precession periods can be compared visually), but it is more than twice as long in the middle panel, demonstrating that ϖ is truly static in this case, on time-scales much longer than the precession periods seen in the other cases. The alignment for the $e_B = 0.4$ case is also associated with smaller disk eccentricity (by a factor of ~ 2) at $r \sim (3 - 10)a_B$.

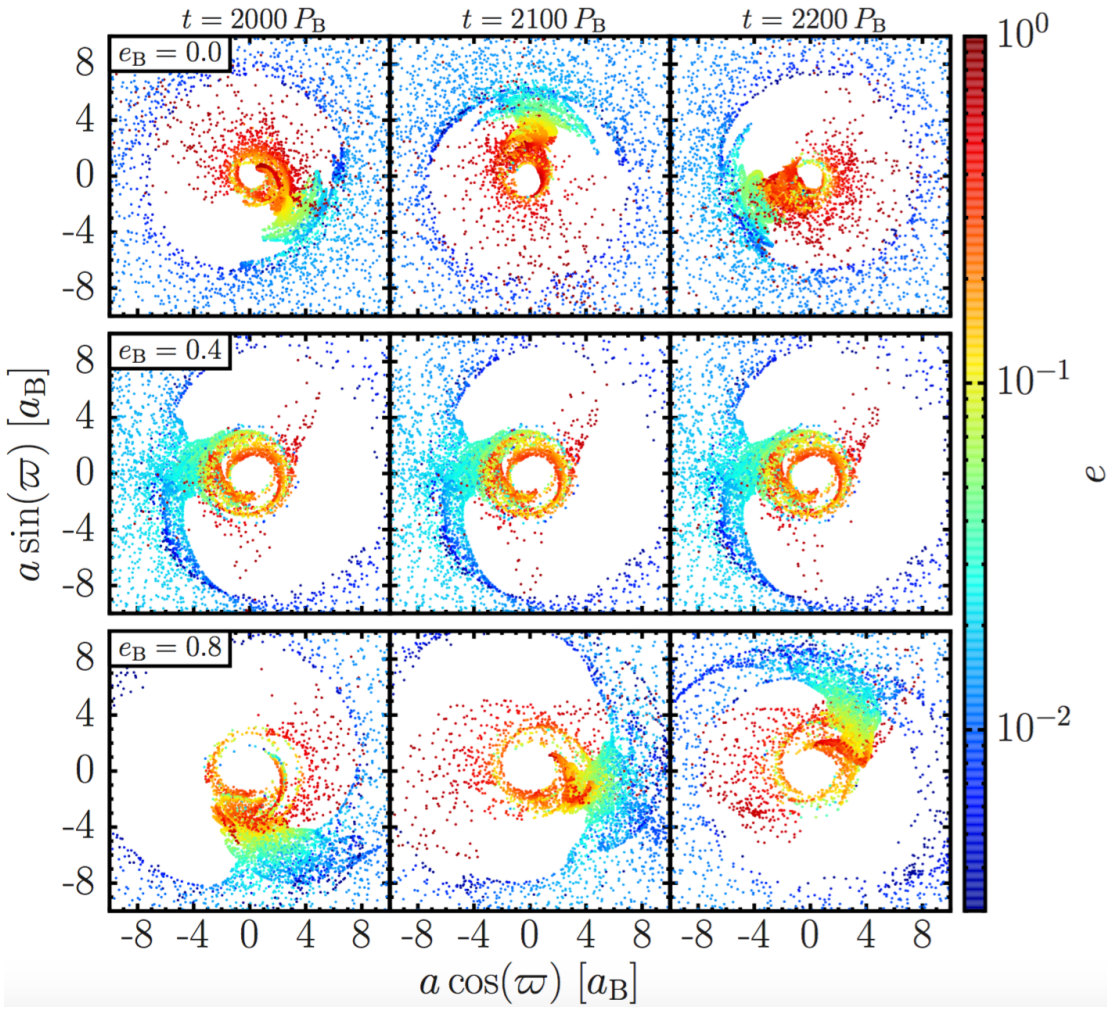

Figure 6.10: Orbital elements of the fluid elements (grid cells) in the disk, only 5 per cent of which are shown, at several points in time, for equal mass binaries with $e_B = 0.0$ (top row), 0.4 (middle row), and 0.8 (bottom row). The radial distance of a point from the origin indicates its semi-major axis, its polar angle represents its pericentre orientation (where the binary line of apses is along the x-axis), and its colour indicates its eccentricity. In all three cases, there are three distinct populations of fluid elements: (i) those with nearly circular orbits ($e \lesssim 0.02$), large semi-major axes ($a \gtrsim 6a_B$), and no preferred pericentre orientation, corresponding to a circular disk; (ii) those with moderate eccentricites ($0.03 \lesssim e \lesssim 0.15$) and semi-major axes between $3a_B$ and $6a_B$, with pericentres clustered around one particular direction at a given time, representing a coherently eccentric part of the disk which precesses (top and bottom rows), or remains locked with the binary (middle row); and (iii) those with very high eccentricity ($e \approx 1$) and $a \lesssim 3a_B$, which represent material plunging toward the binary on nearly radial orbits (accretion streams). For $e_B = 0.0$ and 0.8, the streams are nearly isotropic in ϖ, while for $e_B = 0.4$, they appear to have two preferred orientations that remain fixed in time.

eccentric orbit is characterized by the $m = 1$ pattern in u_r, which may precess (rotate) at a frequency much slower than the local orbital frequency. The power spectrum of the same Fourier component of another fluid variable, such as Σ, would also reveal the same periodicity (due to mass continuity), but we choose u_r since its equilibrium value is nearly zero (except for a small viscous drift). We see that for the $e_{\rm B} = 0.0$ and $e_{\rm B} = 0.8$ cases, the power spectrum is concentrated in a feature with a frequency of a few thousandths of $\Omega_{\rm B}$, which is coherent over a large range of radii. Weaker features at harmonics of this fundamental frequency can also be seen. Also shown in Fig. 6.8 is the apsidal precession rate of a test particle (on a nearly circular orbit) due to the binary quadrupole potential (e.g., Liu et al. 2015a, Eq. 20),

$$\dot{\varpi}_{\rm Q} = \frac{3}{4}\frac{q_{\rm B}}{(1+q_{\rm B})^2}\left(1+\frac{3}{2}e_{\rm B}^2\right)\left(\frac{r}{a_{\rm B}}\right)^{-7/2}\Omega_{\rm B}. \tag{6.9}$$

The features in the power spectrum for $e_{\rm B} = 0$ and $e_{\rm B} = 0.8$ have frequencies close to $\dot{\varpi}_{\rm Q}$ evaluated in the vicinity of $r_{\rm peak}$. Given the steep radial dependence ($r^{-7/2}$) of $\dot{\varpi}_{\rm Q}$, the exact value of r at which the two quantities are equal is not particularly meaningful. However, the fact that the frequency of the feature is similar to the range of $\dot{\varpi}_{\rm Q}$ near the inner edge is significant, since it indicates that the feature is associated with the coherent precession of the inner disk. For $e_{\rm B} = 0.4$ (the middle panel of Fig. 6.8), this feature is conspicuously absent, indicating that the long-term dynamics of the disk are different in this case.

We investigate the eccentricity dynamics further by treating each grid cell as a test particle orbiting in the potential of the binary and, converting its instantaneous position (r, ϕ) and velocity (u_r, u_ϕ) into orbital elements (a, e, ϖ), using energy conservation (the "vis-viva equation"),

$$a(r,\phi,t) = \left(\frac{2}{r} - \frac{u^2}{GM_{\rm B}}\right)^{-1}, \tag{6.10}$$

and the (Runge-Lenz) eccentricity vector,

$$\mathbf{e}(r,\phi,t) = \frac{\mathbf{u}^2\mathbf{r} - (\mathbf{u}\cdot\mathbf{r})\mathbf{u}}{GM_{\rm B}} - \hat{\mathbf{r}}, \tag{6.11}$$

where $\mathbf{e} = [e\cos(\varpi), e\sin(\varpi)]$. We then use the mass-weighted azimuthal average of the eccentricity vector,

$$\langle\mathbf{e}\rangle_\phi(r,t) = \frac{\oint \Sigma(r,\phi,t)\mathbf{e}(r,\phi,t)\mathrm{d}\phi}{\oint \Sigma(r,\phi,t)\mathrm{d}\phi}, \tag{6.12}$$

to define the average eccentricity e and pericentre angle ϖ at each r, as a function of t. The result is shown in Fig. 6.9, for simulations with three different binary eccentricities.

The eccentricity dynamics falls into one of two regimes. In the first regime ($e_{\rm B} = 0.0$ and 0.8 in Fig. 6.9), the eccentric portion of the disk undergoes coherent apsidal precession. The precession periods can be identified with those shown in Fig. 6.8. In the second regime ($e_{\rm B} = 0.4$), the eccentric disk keeps its line of apsides aligned with that of the binary. The distinction between the two regimes is unambiguous. In the second regime, not even a very slow pericentre advance, on the time-scale of the entire simulation (8000 orbits for the case shown in Fig. 6.9), is seen. Instead, the disk pericentre stays truly aligned with that of the binary indefinitely. Apsidal alignment of the disk and binary was also reported by Lubow & Artymowicz (2000) and Pierens & Nelson (2007), although only for an unequal mass binaries. To our knowledge, apsidal alignment with an equal-mass binary–which does not have an octupole potential and cannot excite eccentricity in the secular regime (see Section 6.6.1)–has not been seen in numerical simulations before. Table 6.1 indicates whether the disk precesses or remains aligned with the binary, as well as the precession period, where applicable, for each simulation.

The distinction between the precessing and aligned regimes is further highlighted in Fig. 6.10, which shows snapshots of fluid elements in the disk in the (a, e, ω) phase space at several points in time, for the same cases as shown in Fig. 6.9. In all three cases, the presence of a coherently eccentric inner disk manifests as the clustering of fluid elements in ϖ. In the precessing case, the cluster can be seen to rotate (by about 2/3 and 1/3 of a full rotation, for $e_{\rm B} = 0.0$ and 0.8), over the $200P_{\rm B}$ interval shown, while in the aligned case, the cluster remains at the same ϖ as time progresses. In this figure, there is a distinction between two different populations of fluid elements, one with $e \sim 0.01 - 0.2$ (what we call the eccentric disk proper, between about 3 and $6a_{\rm B}$, represented by yellow/green dots), and another with $e \sim 1$ (red dots), which corresponds to streams penetrating the inner cavity on nearly radial orbits. In the precessing case, the streams do not have of a preferred orientation in ϖ, while for the aligned case, there are two strongly preferred directions (both $\sim 45^\circ$ from the binary line of apsides). This explains the apparent discontinuity in the average ϖ of the disk close to the binary ($r/a_{\rm B} \lesssim 2$) seen in Fig. 6.9: it is an artefact of the accretion streams, rather than of the eccentric disk proper, which does not exist close to the binary. The restricted orientation of the streams in the aligned case ($e_{\rm B} = 0.4$) may be associated with the reduced rate at which angular momentum is transferred to the binary (see Section 6.7).

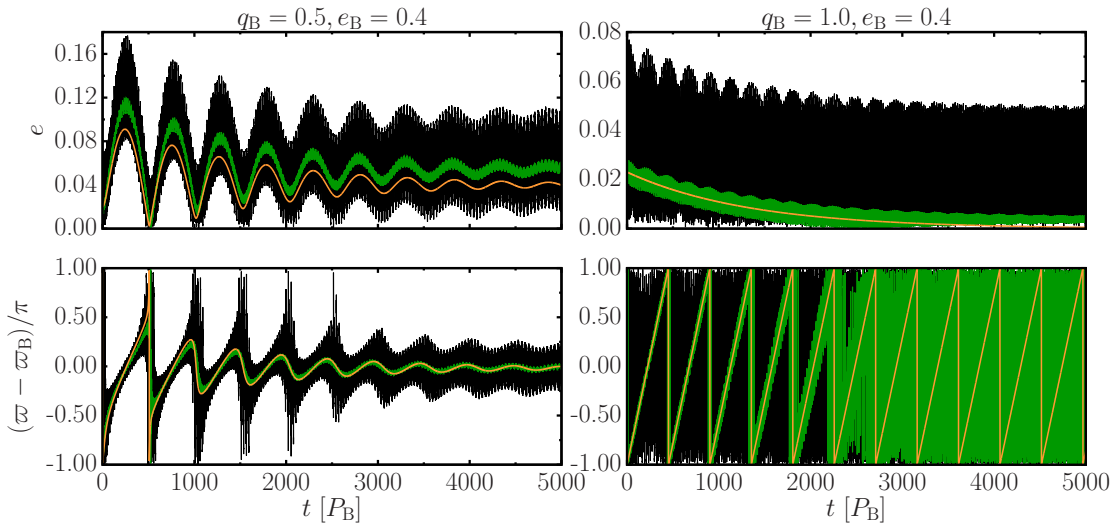

Figure 6.11: Eccentricity (top) and pericentre longitude (bottom) evolution for a test particle with semi-major axis $a = 3.5a_B$ orbiting an eccentric binary with $e_B = 0.4$ (see Section 6.6.1). Two binary mass ratios are shown, $q_B = 0.5$ (left), and 1.0 (right). The orange lines are the results of integrations of the secular equations of motion (Eqs. 6.13, 6.14, and 6.16). The black lines are the results of integrations of the non-secular equation of motion (Eqs. 6.17 and 6.18), and the green lines indicate the e and ϖ of the running time average of the eccentricity vector (from the same integrations), taken over the interval $(t - 5P_B, t + 5P_B)$ for each t. The running average filters out short term variation in order to assess whether ϖ is circulating or librating. For $q_B = 0.5$, the balance of the eccentricity excitation by the binary octupole potential and viscous damping (modeled by Eq. 6.16 or Eq. 6.18) results in the test particle settling into a finite eccentricity orbit with its pericentre aligned with that of the binary. For $q_B = 1.0$, the lack of eccentricity excitation by an octupole potential results in pure damping of the eccenticity: the test particle evolves towards a circular orbit, while ϖ continues to circulate indefinitely (although becoming increasingly less well-defined due to the very small e). The same behavior is seen for both the secular equations and non-secular equations, indicating that short time-scale dynamical forcings from the binary do not modify the long-term evolution of the particle.

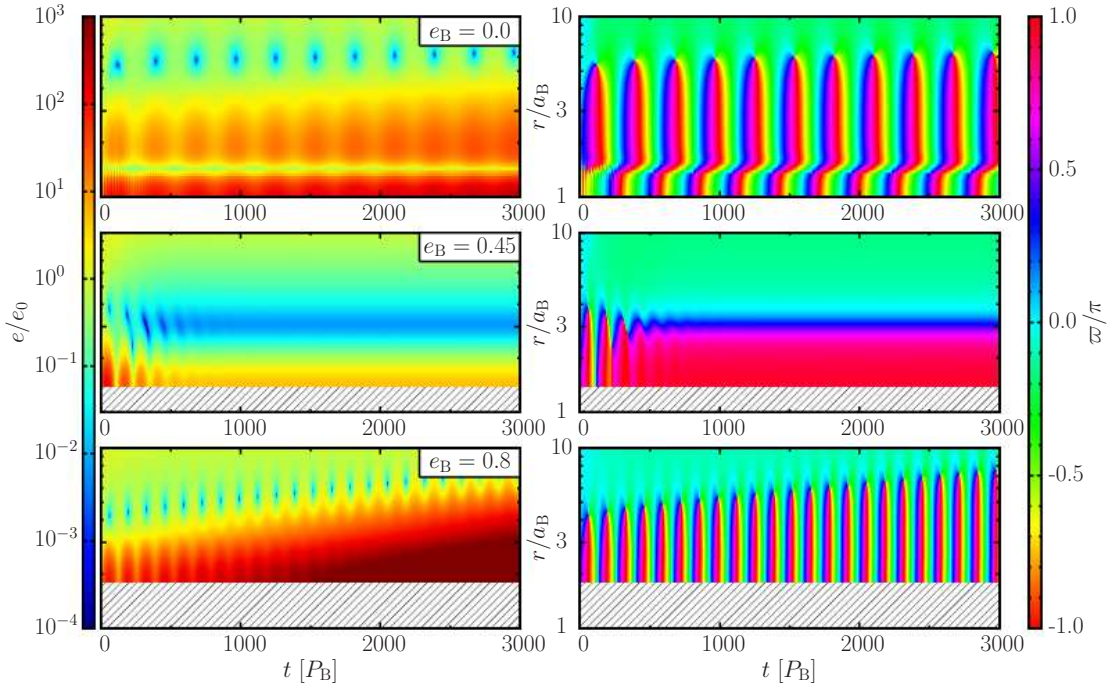

Figure 6.12: Disk eccentricity and longitude of pericentre, as a function of time and radius (as in Fig. 6.9), as calculated from the one dimensional linear fluid eccentricity equation (Eq. 6.19), with eccentricity driven at two eccentric Lindblad resonances (ELRs; see Section 6.6.2). Three different binary eccentricities are shown. In all cases, the binary has $q_B = 1.0$, and the disk has $h = 0.1$ and $\alpha_b = 0.1$ (as in our numerical simulations, except that the shear viscosity parameter α is replaced by a bulk viscosity parameter α_b). For $e_B = 0.0$ (top), the disk eccentricity grows as a result of driving at the $\Omega = \Omega_B/2$ ELR, and the disk precesses. Similarly, for $e_B = 0.8$ (bottom), eccentricity grows as a result of driving at the 1/4 ELR, and the inner disk $[(1.5 - 6)a_B]$ precesses coherently. For $e_B = 0.45$, the eccentricity drivings at both ELRs are too weak to overcome viscous damping, so no eccentricity growth is observed, and the propagation of the eccentricity wave is damped, resulting in a halting of apsidal precession. This behavior may explain the similar qualitative trend seen in our numerical simulations, shown in Fig. 6.9.

6.6 Theoretical Explanations for Disk Eccentricity Excitation, Precession, and Apsidal Locking

In Section 6.5, we presented numerical results demonstrating that the inner region of a circumbinary disk generally becomes eccentric and evolves coherently. We showed that, for both low and high values of $e_{\rm B}$, the pericentre of the disk precesses coherently around the binary. However, for intermediate values of $e_{\rm B}$, the disk instead becomes apsidally locked with the binary. This result is puzzling and unexpected, and so in this section, we explore several possible theoretical explanations for this behavior. This section is self-contained, and can be skipped if the reader wishes to continue on to the presentation of the rest of our numerical results, which resumes in Section 6.7.

6.6.1 Test Particle Dynamics

We first take the simplest approach, considering the dynamics of a test particle which orbits the binary, and which is subject to the tidal potential of the binary as well as a parametrized "frictional" eccentricity damping force.

Secular Dynamics

The long-term dynamics of the particle may be described using the secular approximation, in which the orbital motion of both the binary and the particle are averaged out. The resulting orbital evolution of a free particle, to octupole

order, is given by (e.g., Liu et al. 2015b[2])

$$\frac{1}{\Omega_{\rm K}}\frac{{\rm d}e}{{\rm d}t} = -\frac{15}{64}\frac{q_{\rm B}(1-q_{\rm B})}{(1+q_{\rm B})^3}\left(\frac{a_{\rm B}}{a}\right)^3 \frac{e_{\rm B}(4+3e_{\rm B}^2)}{(1-e^2)^2}\sin(\varpi-\varpi_{\rm B}), \tag{6.13}$$

$$\begin{aligned}\frac{1}{\Omega_{\rm K}}\frac{{\rm d}\varpi}{{\rm d}t} &= \frac{3}{8}\frac{q_{\rm B}}{(1+q_{\rm B})^2}\left(\frac{a_{\rm B}}{a}\right)^2 \frac{(2+3e_{\rm B}^2)}{(1-e^2)^2} \\ &\quad - \frac{15}{64}\frac{q_{\rm B}(1-q_{\rm B})}{(1+q_{\rm B})^3}\left(\frac{a_{\rm B}}{a}\right)^3 \frac{e_{\rm B}(4+3e_{\rm B}^2)(1+4e^2)}{e(1-e^2)^3} \\ &\quad \times\cos(\varpi-\varpi_{\rm B}).\end{aligned} \tag{6.14}$$

In general, the particle and the binary exchange angular momentum (although the amount gained or lost by the binary is negligible in the test particle limit), resulting in oscillations of the eccentricity of the particle, which in the linear regime ($e, e_{\rm B} \ll 1$), has an amplitude of

$$e_{\rm forced} = \frac{5}{8}\left(\frac{1-q_{\rm B}}{1+q_{\rm B}}\right)\frac{a_{\rm B}}{a}\left(\frac{4+3e_{\rm B}^2}{2+3e_{\rm B}^2}\right)e_{\rm B} \tag{6.15}$$

(e.g., Moriwaki & Nakagawa 2004). To model the effect of eccentricity damping (e.g., due to viscosity), we add the following term to the right hand side of Eq. (6.13),

$$\dot{e}_{\rm damp} = -\frac{e}{t_{\rm d,s}}, \tag{6.16}$$

where $t_{\rm d,s}$ is the eccentricity damping time-scale. In the presence of eccentricity damping, the particle can evolve towards a fixed state in which $\varpi - \varpi_{\rm B} = 0$ and $e \approx e_{\rm forced}$ (e.g., Wu & Goldreich 2002). This apsidal alignment can occur only if $e_{\rm forced} \neq 0$, which requires that $q_{\rm B} < 1$ and $e_{\rm B} > 0$ (i.e., the binary octupole potential is non-zero), otherwise the eccentricity of the test particle will continually precess as its eccentricity asymptotically approaches zero.

Figure 6.11 shows several examples of the secular evolution of a test particle subject to eccentricity damping. In these examples, the semi-major axis of the

[2]We have corrected typos in the relevant equations of Liu et al. (2015b): in Eq. (12), $(1-e_2^2)^{5/2}$ should be $(1-e_2^2)^2$, and in Eq. (14), $(4+3e_1)$ should be $(4+3e_1^2)$.

particle is chosen to be $a = 3.5a_{\rm B}$, and the eccentricity damping time is $t_{\rm d,s} = 15/\dot{\varpi}_{\rm Q}$ (see Eq. 6.9). For the unequal mass binary, the argument of pericentre of the test particle undergoes several precession cycles, before librating around that of the binary with a decreasing amplitude, until it is essentially aligned with the binary. The eccentricity of the particle oscillates and eventually approaches a fixed value of $0.04 \approx e_{\rm forced}$. For the equal-mass binary, the eccentricity simply approaches zero, while the argument of pericentre precesses indefinitely.

In this secular framework, an unequal mass binary is required to produce apsidal alignment. This may explain the alignment seen in one of our numerical simulations, with $(q_{\rm B}, e_{\rm B}, \alpha) = (0.2, 0.4, 0.1)$. However, it cannot explain the alignment seen in several of our simulations with equal mass binaries.

Non-Secular Dynamics

Since secular theory cannot produce apsidal alignment without a non-zero octupole potential, we now consider the non-secular (i.e., non-orbit-averaged) dynamics of a test particle. This is potentially useful, since the fluid elements of the eccentric disk have semi-major axes which are only a few times that of the binary ($a/a_{\rm B} \approx 3 - 6$), so the short-term (orbital time-scale) forcings from the binary are not necessarily negligible compared to the long-term, secular forcings. The short-term forcings may give rise to additional eccentricity excitation which is not captured in the secular approximation.

The equation of motion of a test particle (with position vector $\mathbf{r}$) reads

$$\ddot{\mathbf{r}} = -\nabla\Phi + \mathbf{f}_{\rm d}, \tag{6.17}$$

where Φ is the (time-dependent) gravitational potential of the central binary. We

take the damping force $\mathbf{f}_{\rm d}$ to be of the form

$$\mathbf{f}_{\rm d} = -\frac{1}{t_{\rm d,ns}}\left[3(\hat{\mathbf{r}}\cdot\dot{\mathbf{r}})\hat{\mathbf{r}} + (\hat{\mathbf{r}}\times\dot{\mathbf{r}} - r\boldsymbol{\Omega}_{\rm p})\times\hat{\mathbf{r}}\right], \tag{6.18}$$

with $\boldsymbol{\Omega}_{\rm p} = (|\dot{\mathbf{r}}|/|\mathbf{r}|)\hat{\mathbf{z}}$ (e.g., Mardling & Lin 2002). This form of $\mathbf{f}_{\rm d}$ ensures that a circular orbit does not decay. We find that choosing $t_{\rm d,ns} = 2.5t_{\rm d,s}$ results in the same eccentricity damping rates between the secular and non-secular equations. Thus we adopt $t_{\rm d,ns} = 37.5/\dot{\varpi}_{\rm Q}$ in the following numerical examples.

Figure 6.11 shows some numerical results based on integrating Eq. (6.17). The particle starts at $3.5a_{\rm B}$, with a small eccentricity. The eccentricity and argument of pericentre exhibit large amplitude, short-time-scale fluctuations, but on average are very similar to the corresponding secular results. In particular, for a non-equal mass binary, the particle becomes apsidally aligned with the binary, with a finite eccentricity, although still exhibiting fast, order unity fluctuations. The average value of the eccentricity is slightly larger than in the secular calculation due to these fluctuations. For the equal mass binary, the orbital evolution of the test particle is also very similar to the secular calculation. Despite the large amplitude, fast fluctuations, the initial eccentricity of the particle is smoothly damped, and its argument of pericentre continues to precess. Although the average eccentricity becomes vanishingly small, it can still be as large as 0.05 instantaneously, indicating that there is some some additional eccentricity excitation by the non-secular forcings from the binary. However, there is no sign of apsidal alignment for the equal mass binary. We conclude that the additional short-term forcings cannot explain the apsidal alignment of a disk around an equal mass binary.

We note that the results presented in Fig. 6.11 do not include the effect of mean motion resonances, which likely plays an important role in eccentricity

excitation of the disk. In the following subsection, we consider the dynamics of a fluid disk, including resonant eccentricity excitation.

6.6.2 Linear Fluid Dynamics of Eccentric Disks

The dynamics of an eccentric fluid disk can be formally described using linear perturbation theory, in which eccentricity propagates as a slow, one-armed ($m = 1$) density wave. The equation describing the evolution of the complex eccentricity vector $E = e\exp(\mathrm{i}\varpi)$ is (Goodchild & Ogilvie 2006)

$$\begin{aligned} 2r\Omega\frac{\partial E}{\partial t} = &-\frac{\mathrm{i}E}{r}\frac{\partial}{\partial r}\left(r^2\frac{\partial \Phi_2}{\partial r}\right) + \frac{\mathrm{i}E}{\Sigma}\frac{\partial P}{\partial r} \\ &+ \frac{\mathrm{i}}{r^2\Sigma}\frac{\partial}{\partial r}\left[(1-\mathrm{i}\alpha_\mathrm{b})Pr^3\frac{\partial E}{\partial r}\right] \\ &+ \sum_i 2a_\mathrm{B}\gamma_i r\Omega E\delta(r-r_{\mathrm{res},i}), \end{aligned} \tag{6.19}$$

where

$$\Phi_2 = -\frac{GM_\mathrm{B}}{4a_\mathrm{B}}\frac{q_\mathrm{B}}{(1+q_\mathrm{B})^2}\left(1+\frac{3}{2}e_\mathrm{B}^2\right)\left(\frac{r}{a_\mathrm{B}}\right)^{-3} \tag{6.20}$$

is the (time-independent) quadrupole component of the binary potential. The first two terms on the right-hand side of Eq. (6.19) describe precession due to the potential and gas pressure, respectively, and the third term describes the diffusion of eccentricity through the disk (with α_b characterizing the disk viscosity). The fourth term describes the growth of eccentricity at various resonances, each having a growth rate γ_i, and which are idealized as infinitely narrow (i.e., δ functions) in r.

The binary potential can be decomposed into Fourier components,

$$\Phi(r,\phi,t) = \sum_{m,N}\Phi_{m,N}\cos(m\phi - N\Omega_\mathrm{B}t), \tag{6.21}$$

where $\Phi_{m,N}$ is the strength of the potential component having azimuthal number m, rotating with pattern frequency $\omega_{\rm P} = N\Omega_{\rm B}/m$. Eccentricity can be excited in the disk at eccentric Lindblad resonances (ELRs), whose locations are determined by the criterion

$$\Omega = \frac{m\omega_{\rm P}}{m+2}, \tag{6.22}$$

and which have growth rates $\gamma_{m,N}$ that are proportional to $\Phi^2_{m,N}$ (Lubow 1991). We specialize to equal-mass binaries, which have a vanishing $m = 1$ potential component, and for which the 1:3 ELR (e.g., Lubow 1991; Papaloizou et al. 2001; Fleming & Quinn 2017) does not exist. Thus, resonances with $m = 2$ are the strongest, since $\Phi_{m,N} \propto r^{-m-1}$. For a circular binary, only the $\Phi_{m,m}$'s are non-zero, i.e., the only possible pattern frequency is $\omega_{\rm P} = \Omega_{\rm B}$. Thus, the most important resonance corresponds to the $\Omega = \Omega_{\rm B}/2$ commensurability (located at $r = 1.59a_{\rm B}$), and its growth rate is proportional to $\Phi^2_{2,2}$. Resonances with higher values of m are not important, as they are located very close to the binary, where the flow deviates strongly from Keplerian. For eccentric binaries, resonances with different pattern frequencies exist. Those with $\omega_{\rm P} > \Omega_{\rm B}$ are not relevant, since they are located inside the disk truncation radius. However, those with $\omega_{\rm P} < \Omega_{\rm B}$ are located at larger radii, and can be important. The second most important resonance is associated with $\omega_{\rm P} = \Omega_{\rm B}/2$, which is located at the $\Omega = \Omega_{\rm B}/4$ commensurability ($r = 2.52a_{\rm B}$) and has a growth rate proportional $\Phi^2_{2,1}$.

The strengths of the $m = 2$ potential components, to quadrupole order, evaluated at the ELRs, are

$$\Phi_{2,N}(r_{\rm ELR}) \approx -\frac{3N^2}{64} C_N \frac{GM_{\rm B}}{a_{\rm B}} \frac{q_{\rm B}}{(1+q_{\rm B})^2}, \tag{6.23}$$

where the relevant coefficients, to order $e^2_{\rm B}$, are given by $C_2 \approx 1 - 5e^2_{\rm B}/2$ and $C_1 \approx -3e_{\rm B}$ (note that C_N corresponds to the notation $C^{\rm CB}_{2,2,N-2}$ used in Miranda &

Lai 2015). For simplicity, we choose

$$\gamma_0 \equiv \gamma_{2,2}\Big|_{e_B=0} = 2\frac{q_B^2}{(1+q_B)^4}\Omega_B \tag{6.24}$$

as an ansatz for the normalization of the ELR growth rates, using the result of Lubow (1991), where we have set the factors related to the disk geometry and resonance width to unity. We then have

$$\gamma_{2,2} = \left(1 - 5e_B^2\right)\gamma_0, \tag{6.25}$$

for $e_B < (1/5)^{1/2}$ (otherwise $\gamma_{2,2} = 0$), and similarly,

$$\gamma_{2,1} = \frac{9}{16}e_B^2\gamma_0. \tag{6.26}$$

We investigate the dynamics of eccentric disks driven by these two ELRs by integrating Eq. (6.19), discretized on a uniformly spaced grid (with $\Delta r = 0.05a_B$), using a fourth-order Runge-Kutta method. As in our simulations, we adopt the locally isothermal equation of state, $c_s(r) = 0.1r\Omega_K$, viscosity parameter $\alpha_b = 0.1$, and choose the inner boundary to be $r_{in} = (1 + e_B)a_B$. The outer boundary is at $10a_B$. Zero gradient conditions are imposed in the complex eccentricity ($\partial E/\partial r = 0$) at both boundaries, and a damping zone, which relaxes E to its initial condition on the orbital time-scale of the outer boundary, is imposed in the outer 25 per cent of the domain in order to mimic an outgoing wave boundary condition. The background surface density is adapted from the time averaged, azimuthally averaged profiles shown in Fig. 6.2, for the appropriate binary parameters.

The results of our one dimensional, eccentric fluid disk experiments are shown in Fig. 6.12. For $e_B = 0$, the disk eccentricity grows with time, and its pericentre precesses coherently. The precession period is similar to the one seen

in the numerical simulations. A kink feature in ϖ at the location of the inner resonance, similar to the one seen in the full numerical simulations (Fig. 6.9), is also reproduced. For $e_{\rm B} = 0.45$ (this value is chosen because we found that the disk still precesses for $e_{\rm B} = 0.4$ in our 1D model), the eccentricity does not grow, but rather the initial profile of e is reconfigured into a new equilibrium profile, which does not subsequently change with time. The argument of pericentre of the inner disk can be seen to precess through several cycles, with a shorter period than for $e_{\rm B} = 0$, before eventually settling into a fixed (but r-dependent) orientation. For $e_{\rm B} = 0.8$, the disk eccentricity again grows, and the disk precesses, although with a period that is several times shorter than the one seen in the numerical simulations. Thus, our 1D eccentric disk model based on Eq. (6.19) with two resonant driving terms captures many of the key features seen in the full numerical simulations.

The behavior of our 1D model can be qualitatively understood as follows. In the local (WKB) limit, $E \propto {\rm e}^{{\rm i}\omega t - {\rm i}kr}$, far from the central binary, and far from any resonances, Eq. (6.19) reduces to

$$\omega = -\frac{(1 - {\rm i}\alpha_{\rm b})}{2}\frac{k^2 c_{\rm s}^2}{\Omega}, \tag{6.27}$$

which is the dispersion relation for local spiral density waves, $(\omega - m\Omega)^2 = \Omega^2 + k^2 c_{\rm s}^2$, in the limit of low frequency ($\omega \ll \Omega$), and with $m = 1$, with the addition of viscous damping. If the strength of eccentricity driving at a resonance is not strong enough to overcome viscous damping, then not only will the eccentricity not grow, but the wave will not propagate, meaning that no precession will occur. The strength of the the inner ($\Omega/\Omega_{\rm B} = 1/2$) resonance is a decreasing function of $e_{\rm B}$, while that of the outer resonance ($\Omega/\Omega_{\rm B} = 1/4$) is an increasing function of $e_{\rm B}$. Therefore, for small values of $e_{\rm B}$ (including zero), the inner resonance is strong and can effectively drive eccentricity growth and precession

in the disk. For sufficiently large values of $e_{\rm B}$, the inner resonance is weak (or has shut off entirely), while the outer resonance is strong, and can drive eccentricity growth and precession. For intermediate $e_{\rm B}$, neither resonance is strong, and so the complex eccentricity E freezes out. Thus, our 1D model qualitatively explains two features in our numerical simulations: i) the lack of precession for intermediate values of $e_{\rm B}$, and ii) the associated reduced disk eccentricity, compared to the cases of low or high values of $e_{\rm B}$.

There are several caveats to consider when comparing our 1D eccentric disk models (Fig. 6.12) and full 2D simulations (Fig. 6.9). First, the linear treatment in the 1D model only captures eccentricity growth (which occurs for precessing disks) or non-growth (for non-precessing disks), and cannot capture its saturation, which presumably occurs due to non-linear effects. In our 2D simulations, eccentricity always grows and settles to some equilibrium value (actually a range of values at different locations in the disk), although it is smaller by approximately a factor of 2 in the apsidally aligned regime, compared to in the precessing regime. Second, the argument of pericentre ϖ in Eq. (6.19) is only defined relative to an arbitrary reference angle, rather than with respect to the binary argument of pericentre $\varpi_{\rm B}$. This is a consequence of considering the binary gravitational potential only to quadrupole order (this is appopriate, since here we are focusing on equal mass binaries, for which the octupole potential vanishes). As a result, when precession of the disk is halted in our 1D model (the middle panel of Fig. 6.12), the value of ϖ at which it freezes out is also arbitrary. Therefore, true alignment of the disk pericentre with that of the binary is not captured in this 1D approach.

6.7 Long-Term Mass Accretion and Net Angular Momentum Transfer

In this section, we use our simulations to determine the long-term net angular momentum transfer rate from the disk to the binary, taking into account mass accretion and gravitational torques. Given the large variabilities of the disk accretion on various time-scales (see Sections 6.4 and 6.5), it is essential that the simulations be carried out over sufficiently long times so that the disk reaches a true quasi-steady state in terms of mass accretion.

6.7.1 Evolution of Global Mass Accretion Rate Profile

Figure 6.13 (top panel) shows examples of both instantaneous radial profiles of $\dot{M}$ (computed using piecewise parabolic reconstruction of the fluid variables), as well as its time-averaged profile. In these examples, the time-averaged profile is computed over an interval of $250P_{\rm B}$ (approximately the disk precession period). We see that the instantaneous $\dot{M}$ profiles exhibit large fluctuations (in both magnitude and sign), with an amplitude $20-60$ times larger than the mass supply rate $\dot{M}_0$. Nonetheless, the average $\dot{M}$ profile is very flat, such that its fluctuations are not visible when plotted on the same scale as the instantaneous profiles. Note that in Fig. 6.13, some of the sharp peaks in the instantaneous profiles are captured by only a few grid points. In our high resolution runs, we find that the amplitude of these features are somewhat reduced, as is the amplitude of those in the averaged profiles. Our standard resolution runs typically achieve an accuracy of about 10 per cent in the averaged $\dot{M}$.

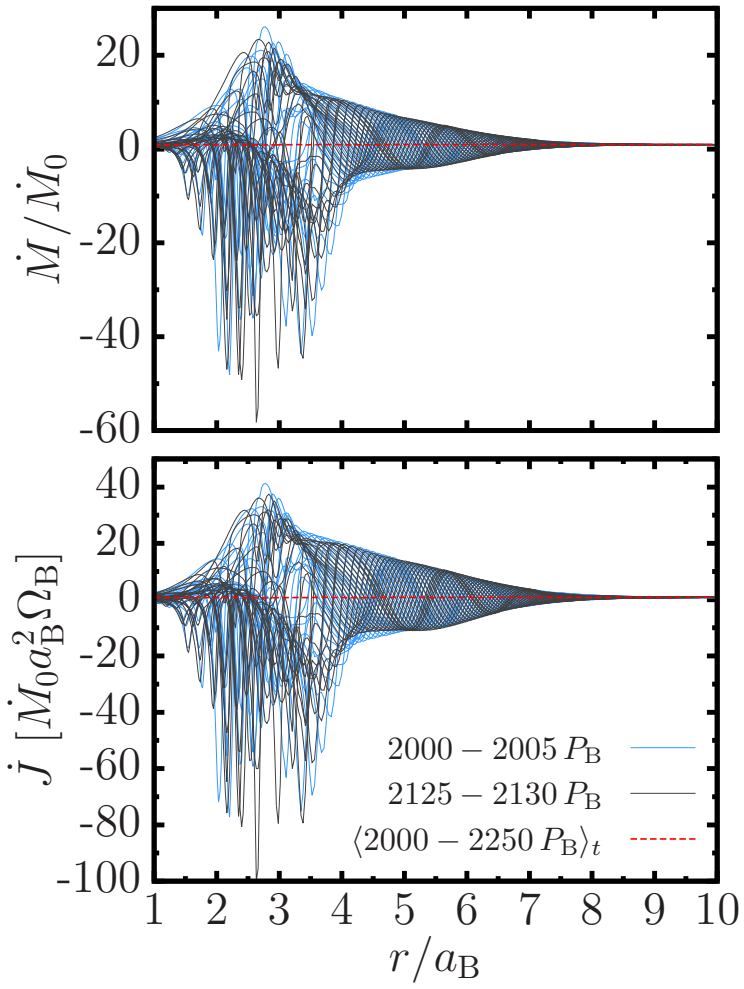

Figure 6.13: Radial profiles of the mass accretion rate (top) and net angular momentum accretion rate (bottom) for the simulation with $q_{\rm B} = 1$, $e_{\rm B} = 0$, and $\alpha = 0.1$. The solid lines are instantaneous profiles, at 50 different times, sampled over two $5P_{\rm B}$ intervals $125P_{\rm B}$ apart. The dashed lines are the time-averaged profiles, taken over $250P_{\rm B}$. Although there are large fluctuations in these quantities both in time and radius, suitable time averaging results in remarkably flat profiles for both.

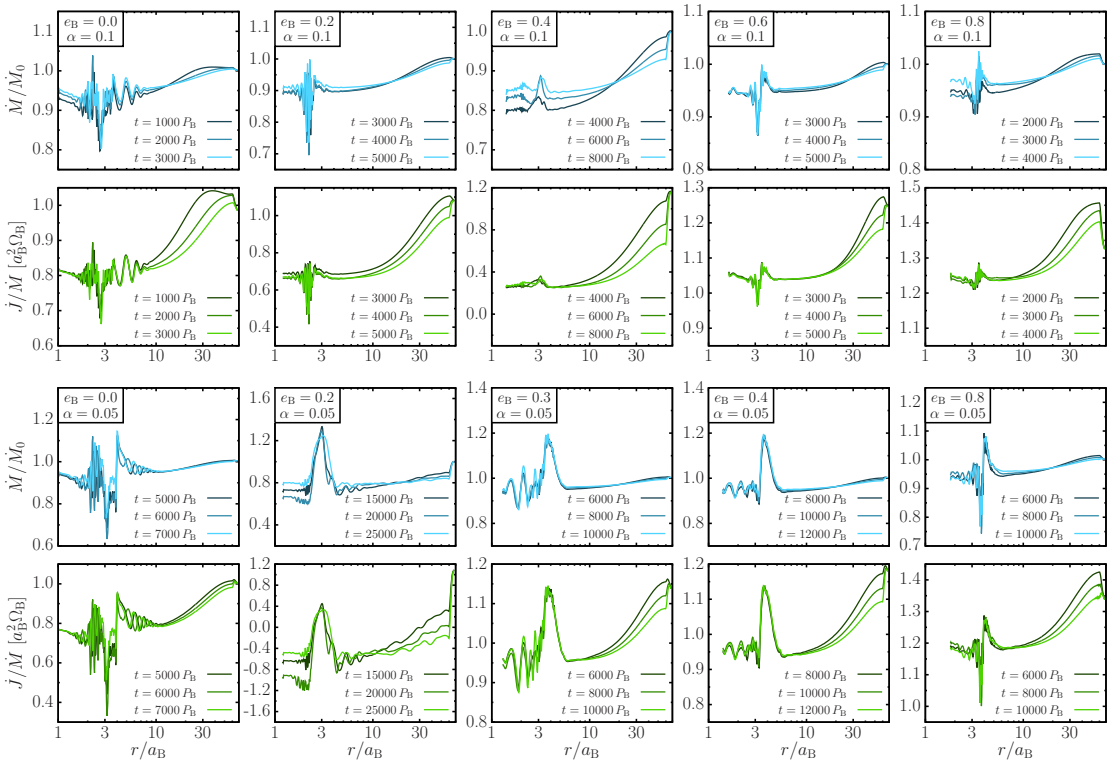

Figure 6.14: Long-term time evolution of the profiles of the mass accretion rate, $\dot{M}$, and the ratio of the angular momentum and mass accretion rates, $\dot{J}/\dot{M}$, for equal mass binaries with different values of e_B and α. Both $\dot{M}$ and $\dot{J}$ profiles are averaged over approximately one precession period [about $(300 - 600)P_B$, for cases in which the disk precesses around the binary], or $50P_B$ (for cases in which the disk is apsidally locked to the binary), starting from the indicated time. The range of the y-axis varies between different panels, in order to focus on the residual radial variations in the profiles of the two quantities after time averaging. Note that the residual variations are much smaller than the instantaneous variations depicted in Fig. 6.13. From the time-averaged profiles, we see that 1) $\dot{M}$ in the inner disk is very close to, but smaller than, the rate supplied at the outer boundary, and slowly approaches it with time, and 2) $\dot{J}/\dot{M}$ approaches a steady value in the inner region of the disk, and the size of this region grows with time. Together, these indicate that an ever-growing region of the inner disk is approaching a self-consistent quasi-steady state.

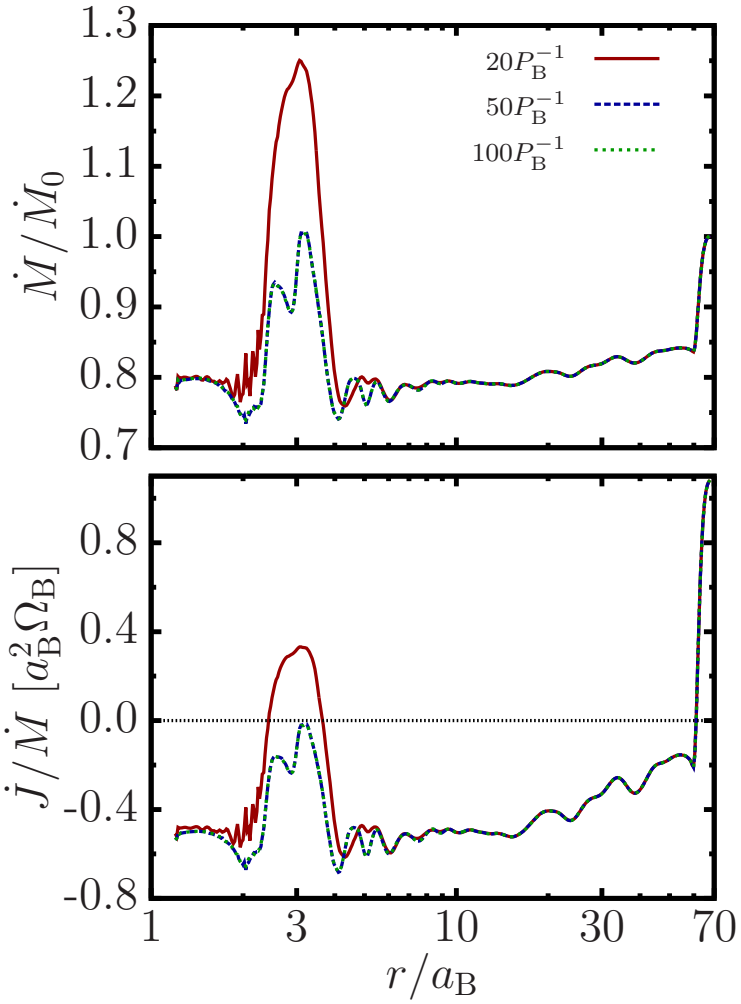

Figure 6.15: Time averaged $\dot{M}$ and $\dot{J}$ for the simulation with $q_{\rm B} = 1.0, e_{\rm B} = 0.2, \alpha = 0.05$, each taken over 50 binary orbits, with several time sampling rates, as indicated. Flatter profiles are obtained by using higher time sampling rates.

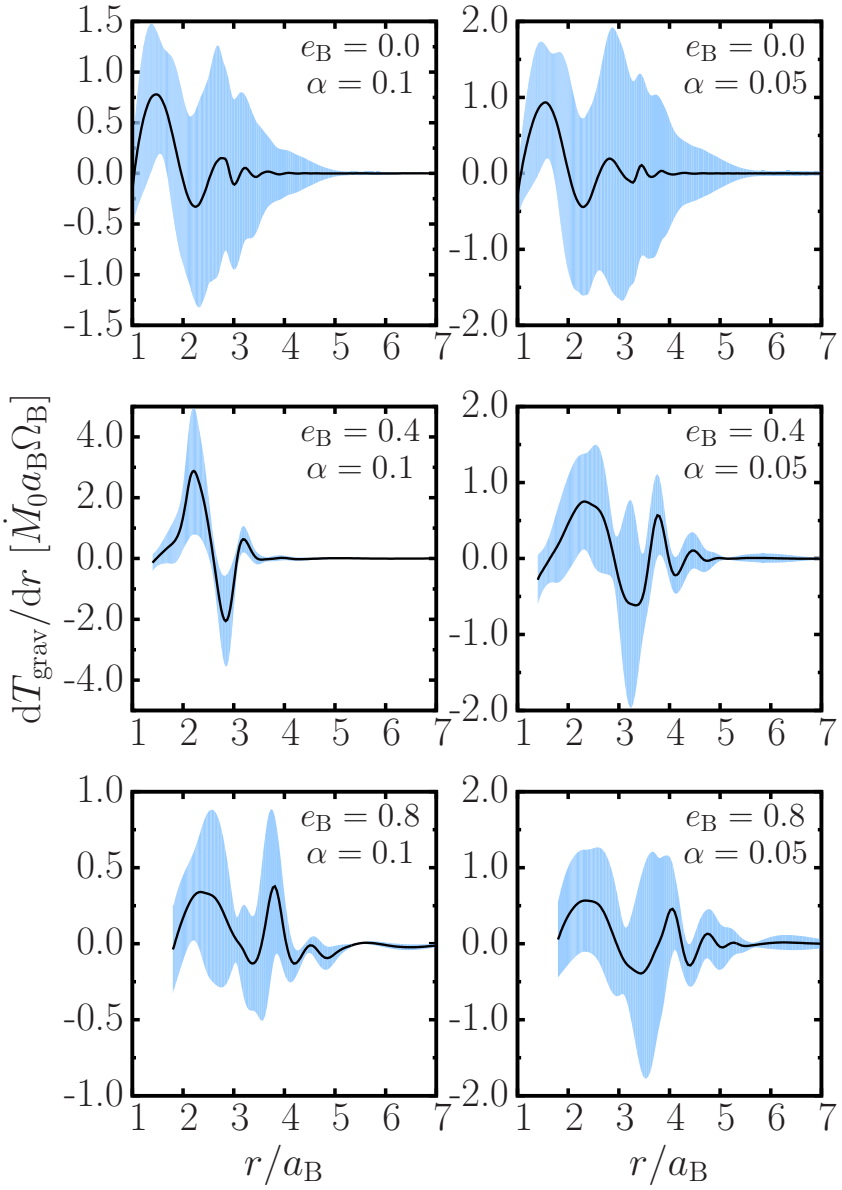

Figure 6.16: Time averaged gravitational torque density dT_{grav}/dr (solid line), and its 1σ variations (shaded region), for equal mass binaries with several different eccentricities and two values of the disk α. The torque density is an oscillatory function of r and can be either positive or negative, and most of the torque is exerted on the disk within a radius of about $6a_B$.

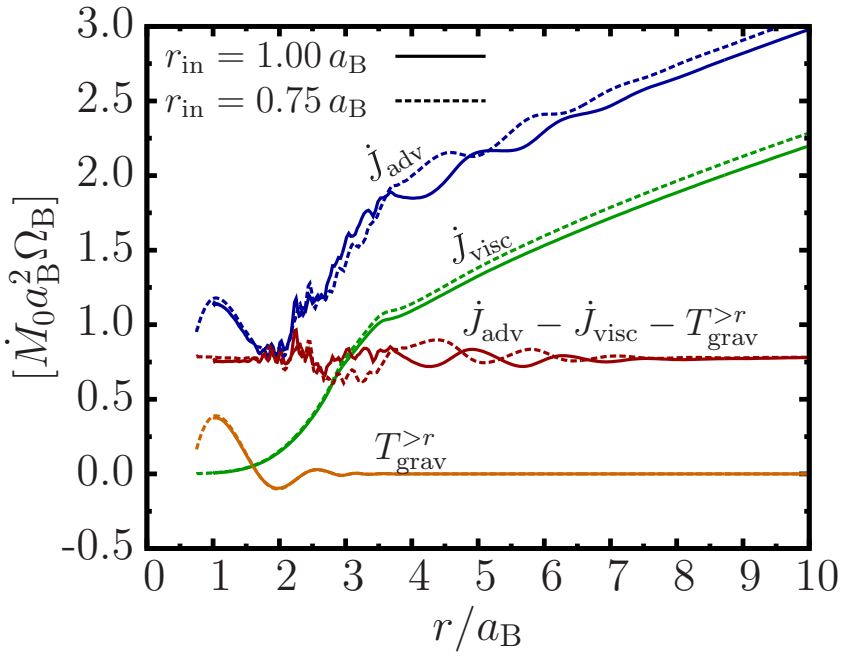

Figure 6.17: Various contributions to the net angular momentum transfer rate as a function of radius for the simulation with $q_{\rm B} = 1$, $e_{\rm B} = 0$, and $\alpha = 0.1$. The net angular momentum transfer includes the advective, viscous, and gravitational contributions (see Eq. 6.28). The results for two different runs are shown: one with the standard inner boundary location, $r_{\rm in} = a_{\rm B}$ (solid lines), and one with an inner boundary located at $r_{\rm in} = 0.75 a_{\rm B}$ (dashed lines). In both runs, all quantities are evaluated at around $2000 P_{\rm B}$, and averaged over $250 P_{\rm B}$.

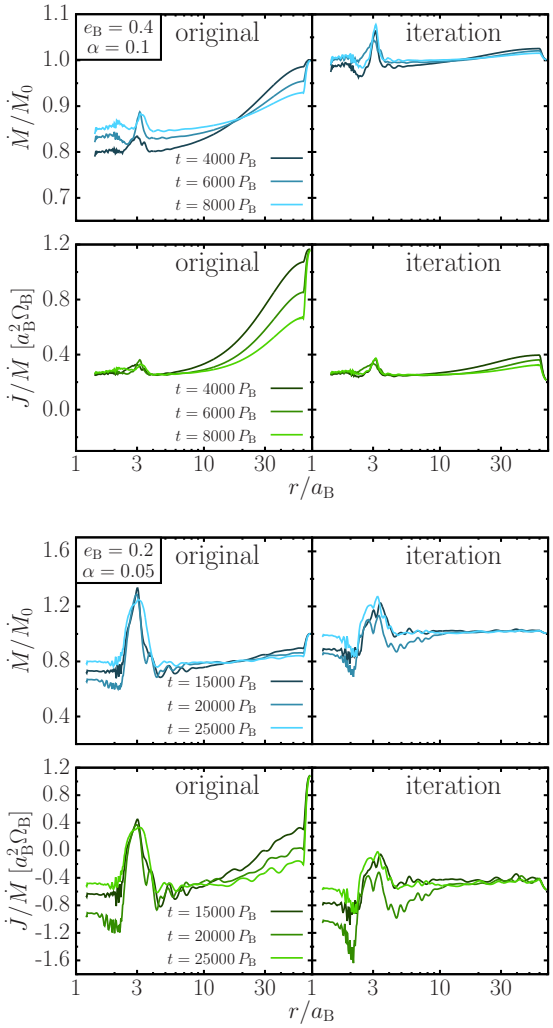

Figure 6.18: Time averaged profiles of $\dot{M}$ and $\dot{J}/\dot{M}$ at several different times, as in Fig. 6.14, for two "unusual" cases (those which result in a minimum values of $\dot{J}/\dot{M}$ at the inner disk; see Fig. 6.19). These are shown for both the original runs (left), which use the standard initial conditions and outer boundary condition (see Eq. 6.3), and for the iterated runs (right), for which the initial conditions and outer boundary condition are prescribed using the value of l_0 determined from the original run. Both are evolved for the same amount of time. In the iterated runs, both the $\dot{M}$ and $\dot{J}/\dot{M}$ profiles become flatter throughout the entire disk compared to the original runs, and the value of $\dot{M}$ in the inner disk is closer to the supply rate. This indicates that the value of $l_0 = \langle \dot{J} \rangle / \langle \dot{M} \rangle$ determined from the original run is close to the true global steady-state value for the disk.

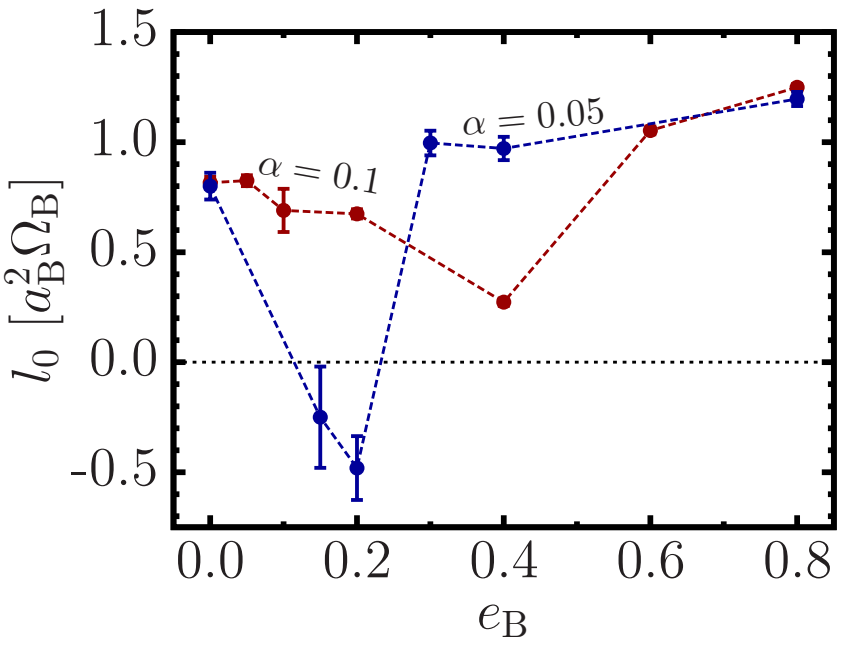

Figure 6.19: Net angular momentum per unit mass received by the binary, $l_0 = \langle \dot{J} \rangle / \langle \dot{M} \rangle$, as a function of $e_{\rm B}$, for accretion onto equal-mass binaries, for two different values of the disk viscosity parameter α. These values are determined by averaging the steady-state profiles of $\dot{J}/\dot{M}$ (see Fig. 6.14) from $r_{\rm in}$ to $10a_{\rm B}$, and the error bars quantify the variations in these profiles. We see that l_0 is positive for most cases, except for when $\alpha = 0.05$ and $e_{\rm B} \sim 0.15 - 0.2$. The dip in the l_0 value at intermediate binary eccentricities (0.4 for the $\alpha = 0.1$ case, and 0.2 for the $\alpha = 0.05$ case) corresponds to the case in which the inner eccentric disk is apsidally aligned with the binary (see Section 6.5).

We now focus on the small residual fluctuations in the time-averaged profiles of $\dot{M}$. Figure 6.14 shows the results for 10 different simulations, all with $q_{\rm B} = 1.0$, and different values of $e_{\rm B}$ and α. The averaging is performed over either approximately one precession period, with a sampling rate of $5/P_{\rm B}$, or over $50P_{\rm B}$, with a sampling rate of $20/P_{\rm B}$, for cases in which the inner disks are apsidally locked. We see that near the end of the simulations, many of the $\langle\dot{M}\rangle$ profiles have residual fluctuations with an amplitude of 30 per cent or less, and in some cases, less than 10 per cent. In the worst case, $(q_{\rm B}, e_{\rm B}, \alpha) = (1.0, 0.2, 0.05)$, the amplitude exceeds 50 per cent. Some of this can be attributed to imperfect averaging, and the amplitude can be further reduced with a higher rate of time sampling. This is demonstrated in Fig. 6.15, in which the sampling rate is increased to $50/P_{\rm B}$, resulting in smaller fluctuations (about 30 per cent in amplitude). No further advantage is gained by increasing the rate to $100/P_{\rm B}$. The $\langle\dot{M}\rangle$ profiles for other cases could also be made flatter with increased an time sampling rate. Thus, the amplitude of the fluctuations in Fig. 6.14 can be considered an upper limit to the real residual fluctuations. We conclude that, at the end of each simulation, $\langle\dot{M}\rangle$ is constant with r to within 30 per cent (or much less), i.e., mass is flowing at an approximately steady rate throughout the disk.

We see from Fig. 6.14 that, in general, at a given time, the time-averaged value of $\dot{M}$ at $r_{\rm in}$ is approximately equal to its value at $\sim 10a_{\rm B}$. This value represents the average mass flow rate through the inner disk onto the binary. In the outer disk, $\dot{M}$ smoothly transitions to its value at $r_{\rm out}$, $\dot{M}_0$. At the end of each simulation, the inner disk $\langle\dot{M}\rangle$ is typically 80 – 90 per cent of $\dot{M}_0$. The small difference between the inner disk $\langle\dot{M}\rangle$ and $\dot{M}_0$ is a consequence of the "incorrect" initial disk density profile: Eqs. 6.3 and 6.4 assume that the specific angular momentum accreted onto the binary is $\sqrt{GM_{\rm B}r_{\rm in}}$. In Section 6.7.2, we show that

starting with better initial conditions (as determined by the angular momentum accretion rate) results in the accretion rate profile $\langle \dot{M} \rangle$ becoming indistinguishable from $\dot{M}_0$ in a much shorter time.

6.7.2 Net Angular Momentum Transfer Rate to the Binary: Calculation and Result

The rate of net angular momentum transfer across the disk, $\dot{J}$, is

$$\dot{J}(r,t) = \dot{J}_{\rm adv} - \dot{J}_{\rm visc} - T^{>r}_{\rm grav} \tag{6.28}$$

(see Appendix D). The different terms contributing to $\dot{J}$ are the rate of change of angular momentum due to advection,

$$\dot{J}_{\rm adv} = -\oint r^2 \Sigma u_r u_\phi {\rm d}\phi, \tag{6.29}$$

the viscous torque,

$$\dot{J}_{\rm visc} = -\oint r^3 \nu \Sigma \left[\frac{\partial}{\partial r}\left(\frac{u_\phi}{r}\right) + \frac{1}{r^2}\frac{\partial u_{\rm r}}{\partial \phi} \right] {\rm d}\phi, \tag{6.30}$$

and the (integrated) gravitational torque,

$$T^{>r}_{\rm grav} = \int_r^{r_{\rm out}} \frac{{\rm d}T_{\rm grav}}{{\rm d}r} {\rm d}r, \tag{6.31}$$

where

$$\frac{{\rm d}T_{\rm grav}}{{\rm d}r} = -\oint r\Sigma \frac{\partial \Phi}{\partial \phi} {\rm d}\phi \tag{6.32}$$

is the gravitational torque density. In general, $\dot{J}$ is a function of r and t. However, in a quasi-steady state state, its time-averaged value, $\langle \dot{J} \rangle$, is independent of r, so that angular momentum flows at a constant rate through the disk and onto the binary. The quasi-steady state is described by the eigenvalue

$$l_0 = \frac{\langle \dot{J} \rangle}{\langle \dot{M} \rangle}, \tag{6.33}$$

which is the angular momentum per unit mass accreted onto the binary. We now examine the evolution of $\langle \dot{J} \rangle$ towards a steady state, in order to compute the value of l_0 for each simulation.

Angular Momentum Transfer Rate Profile

Figure 6.13 (lower panel) shows examples of the instantaneous and the time-averaged profiles of the net angular momentum transfer rate, $\dot{J}$. As in the case of the $\dot{M}$ profiles, the instantaneous fluctuations in $\dot{J}$ are very large (several orders of magnitude larger than the average value), but the fluctuations in the time-averaged profile are much smaller. In order to determine that the disk is evolving towards a quasi-steady state, we must examine these small residual fluctuations in $\dot{J}$.

We first look at the contribution to the angular momentum transfer due to the gravitational torque from the tidal potential of the binary. Although the radially-integrated gravitational torque is the quantity which contributes to $\dot{J}$, it is illustrative to inspect its radial derivative, the gravitational torque density $\mathrm{d}T_{\mathrm{grav}}/\mathrm{d}r$. Examples of its time-averaged profiles and temporal variations are shown in Fig. 6.16. We see that the magnitude of the torque density is only appreciable for $r \lesssim 6a_{\mathrm{B}}$. The torque density is an oscillatory function of r, and can change sign. Thus, it resembles the wave function of an isolated Lindblad resonance (e.g., Meyer-Vernet & Sicardy 1987), or the combination of a few resonances. This type of torque density profile is different from the one arising from the superposition of many Lindblad resonances [e.g., in the impulse approximation, $\mathrm{d}T_{\mathrm{grav}}/\mathrm{d}r \propto 1/(r - a_{\mathrm{B}})^4$; see Lin & Papaloizou 1979], which is only relevant for small mass ratios, and not the mass ratios considered in our simulations

($q_{\rm B} = 0.2 - 1.0$).

We now look at the different contributions to $\dot{J}$. Figure 6.17 shows examples of the time-averaged profiles of the three terms on the right-hand side of Eq. (6.28), which are due to advection, viscous torque, and gravitational torque (the latter being the integral of the quantity shown in Fig. 6.16), as well as the net angular momentum transfer rate, $\dot{J}$, for two different simulations. Both simulations have the same binary and disk parameters, but one uses the standard inner boundary location adopted in the rest of our simulations ($r_{\rm in} = a_{\rm B}$ in this case), while the other uses a slightly smaller inner boundary radius, $r_{\rm in} = 0.75 a_{\rm B}$ (in this simulation, a larger number of radial grid cells were used so that both runs have the same spatial resolution). In both simulations, the profiles of each of the contributions to $\dot{J}$ are very similar, and we find the same nearly constant value of the total $\langle \dot{J} \rangle$ in the inner disk. Thus, an equilibrium value $\langle \dot{J} \rangle$ is reached the inner disk, and it does not depend on the location of the inner boundary.

Figure 6.14 shows the $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ profiles at different times for 10 different simulations. The fluctuations in these averaged profiles are much smaller than in the instantaneous profiles depicted in Fig. 6.13. At the end of each simulation, $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ is nearly constant in the inner part of the disk ($r \lesssim 10 a_{\rm B}$). The fluctuations around the average are typically 25 per cent, but in some cases are less than 10 per cent. In the "worst" case, $(q_{\rm B}, e_{\rm B}, \alpha) = (1.0, 0.2, 0.05)$, the fluctuations are of order unity, although some of this can be attrubuted to the time sampling rate (see Fig. 6.15). For this reason, we ran this simulation for much longer than others ($25000 P_{\rm B}$), in order to ensure relaxation of the inner disk. In all cases, $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ transitions at large radii ($r > 10 a_{\rm B}$) to its prescribed initial value at $r_{\rm out} = 70 a_{\rm B}$, corresponding to $\dot{J}(r_{\rm out}) = \dot{M}_0 \sqrt{G M_{\rm B} r_{\rm in}}$ (see Eq. 6.3).

Iterated Runs

Since $\dot{J}(r_{\rm out})$ is fixed, and since the damping zone near the outer boundary relaxes $\dot{J}$ to this value, $\langle \dot{J} \rangle$ in the outermost part of the disk is never equal to $\langle \dot{J} \rangle(r_{\rm in})$, which is computed self-consistently. If we were to run the simulation for a viscous time-scale at the boundary of the damping zone, $r_{\rm damp} = 60a_{\rm B}$ ($\sim 3 \times 10^4 P_{\rm B}$ for $\alpha = 0.1$), we would see $\langle \dot{J} \rangle$ take on a constant value for $r_{\rm in} < r < r_{\rm damp}$, and transition from this value to $\dot{J}(r_{\rm out})$ in the damping zone. To produce a true steady state, in which $\langle \dot{J} \rangle$ is constant for all r, an iterative process is required. In this process, we use the value of $\langle \dot{J} \rangle$ to construct the initial conditions and outer boundary condition of a new simulation [by replacing $(r_{\rm in}/r)^{1/2}$ in Eq. (6.3) with $l_0/\sqrt{GM_{\rm B}r}$], so that a steady state characterized by l_0 is imposed in the outer disk. Once the inner disk has viscously relaxed, we can achieve a global quasi-steady state, which has $\langle \dot{J} \rangle(r_{\rm in}) \approx \dot{J}(r_{\rm out})$. We performed this iteration for two cases: $(q_{\rm B}, e_{\rm B}, \alpha) = (1.0, 0.4, 0.1)$ and $(1.0, 0.2, 0.05)$. These are cases are of particular interest, as they have the lowest values of l_0 (for their corresponding values of $q_{\rm B}$ and α).

The results of the iterated simulations are shown in Fig. 6.18. In the original runs, $\langle \dot{M} \rangle$ at $r_{\rm in}$ evolves slowly towards $\dot{M}_0$, reaching a value of $\sim 0.8\dot{M}_0$ at the end of the simulation, for both simulations shown. In the iterated runs, $\langle \dot{M} \rangle$ at $r_{\rm in}$ reaches a value very close to $\dot{M}_0$ much sooner, since the initial conditions more closely match the true quasi-steady state of the disk. In the inner disk, $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ has a value close to the one in the original runs, but it is much closer to being constant throughout the entire disk, and the transition to the outer boundary value is less extreme. Thus, the self-consistent angular momentum transfer rate calculated for the the inner disk is independent of the artificially imposed value

at the outer boundary, and represents the true quasi-steady state of the disk. It is therefore not necessary to perform this iteration for all of our simulations, as it does not affect the profile of $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ in the inner disk (i.e., the value of l_0).

Value of the Eigenvalue l_0

We have shown that nearly constant profiles of $\langle \dot{J} \rangle$ and $\langle \dot{M} \rangle$, representing a quasi-steady state, are achieved in the inner disk at the end of our simulations. Thus, we now compute the value of the eigenvalue which characterizes the quasi-steady state, l_0, for each simulation. In a perfect quasi-steady state, in which $\langle \dot{J} \rangle$ and $\langle \dot{M} \rangle$ are independent of r, l_0 is given by Eq. (6.33). Since, in our simulations, $\langle \dot{J} \rangle$ and $\langle \dot{M} \rangle$ are not truly independent of r, we compute the value of l_0 by taking the radial average of $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ between $r_{\rm in}$ and $r_{\rm cut} = 10 a_{\rm B}$,

$$l_0 = \frac{1}{r_{\rm cut} - r_{\rm in}} \int_{r_{\rm in}}^{r_{\rm cut}} \frac{\langle \dot{J} \rangle (r)}{\langle \dot{M} \rangle (r)} {\rm d}r, \tag{6.34}$$

using $\langle \dot{J} \rangle$ and $\langle \dot{M} \rangle$ at the end of each simulation. We also compute the standard deviation of $\langle \dot{J} \rangle / \langle \dot{M} \rangle$ over the same radial interval,

$$\Delta l_0 = \left\{ \frac{1}{r_{\rm cut} - r_{\rm in}} \int_{r_{\rm in}}^{r_{\rm cut}} \left[\frac{\langle \dot{J} \rangle (r)}{\langle \dot{M} \rangle (r)} - l_0 \right]^2 {\rm d}r \right\}^{1/2}, \tag{6.35}$$

in order to quantify the systematic uncertainty in l_0.

The results are given in Table 6.1 and shown in Fig. 6.19 for simulations with equal mass binaries (the values of l_0 for the rest of our simulations are only given in Table 6.1). We see that l_0 almost always has a positive value[3]. Thus,

[3]Figure 6.19 contains an additional point, corresponding to the case $(e_{\rm B}, \alpha) = (0.15, 0.05)$, which is not analysed in detail in the other numerical results sections. This case, which was evolved for $10000 P_{\rm B}$, was found to have a negative value of l_0. It serves to demonstrate that the calculation of l_0 for the case $(e_{\rm B}, \alpha) = (0.2, 0.05)$, which would otherwise be the only occurrence of a negative l_0, is not spurious.

on average, the binary gains net angular momentum. For $e_B = 0$, the value of l_0 (about 0.80, in units of $a_B^2\Omega_B$) is the same to within our uncertainty for both $\alpha = 0.1$ and 0.05. As e_B increases, l_0 first decreases, reaching a minimum (0.27 for $\alpha = 0.1$ and -0.48 for $\alpha = 0.05$), before increasing again. At $e_B = 0.8$, the values of l_0 are almost the same for both values of α (1.25 for $\alpha = 0.1$ and 1.20 for $\alpha = 0.05$). For both values of α, the minimum value of l_0 occurs at the value of e_B for which the disk is apsidally locked with the binary (see Section 6.5).

6.7.3 Net Angular Momentum Transfer Rate to the Binary: Discussion

Figure 6.19 represents the most important result of this paper. We find that the (long-term averaged) specific angular momentum received by the binary, $l_0 = \langle \dot{J} \rangle / \langle \dot{M} \rangle$, is positive for most binary eccentricities (including $e_B = 0$). This directly contradicts the previous numerical results (e.g., MacFadyen & Milosavljević 2008), as well as the the assumption adopted in many papers on the disk-driven merger of supermassive black hole binaries (e.g., Armitage & Natarajan 2002; Haiman et al. 2009; Chang et al. 2010).

As we have emphasized above, l_0 is an eigenvalue of the accretion flow and can only be determined by the global solution of the flow with a proper treatment of the boundaries. In our simulations we have only considered two values of the viscosity parameter ($\alpha = 0.1$ and 0.05) and the disk aspect ratio $H/r = 0.1$. Can a smaller viscosity qualitatively change our result? (For example, a binary surrounded by a "non-accreting" disk would lose angular momentum to the disk through gravitational torque.) We think this is unlikely. Indeed, the angu-

lar momentum flux across the disk can be written as (see Eq. 6.28)

$$\dot{J}(r,t) = \dot{M}l - 3\pi\nu\Sigma l - \Sigma l F(r), \tag{6.36}$$

where we have used the fact that the gravitational torque is proportional to the surface density, and thus $T_{\rm grav}^{>r} = \Sigma l F(r)$, with $F(r) > 0$. In steady state, $\dot{J} = \dot{M}l_0$, and the disk surface density is given by

$$\Sigma = \frac{\dot{M}(1 - l_0/l)}{3\pi\nu + F(r)}. \tag{6.37}$$

Thus, while a reduced viscosity indeed increases the surface density, it does not necessarily change the balance between the advective, viscous, and gravitational angular momentum fluxes.

Among previous numerical studies, the work of MacFadyen & Milosavljević (2008) was the most similar to ours. These authors considered $H/r = 0.1$ but a lower disk viscosity ($\alpha = 0.01$). (They also adopted a polar grid in the domain between $r_{\rm in} = a_{\rm B}$ and $r_{\rm out} = 100a_{\rm B}$.) With such a small viscosity, the "viscous relaxation" radius (see Eq. 6.6) at $t = 4000P_{\rm B}$ (the typical duration of their runs) is only about $3a_{\rm B}$. Moreover, their initial surface density profile is far from the steady state even for $r \gg r_{\rm in}$. Thus, we suggest that their findings concerning the reduction of mass accretion onto the binary and the dominance of the gravitational torque relative to advective torque (therefore a negative l_0) reflected only the "transient" phase of their simulations. Several more recent numerical studies (e.g., D'Orazio et al. 2013; Farris et al. 2014) have explored various aspects of circumbinary accretion, but did not examine the detailed balance of angular momentum transfer.

Shi et al. (2012) obtained a positive value of l_0 in their MHD simulations of circumbinary disks. However, as the duration of their simulations is only

$\sim 100P_{\rm B}$ (due to the costly 3D numerical method), it is unlikely that a quasi-steady state was reached, so their value of l_0 may not properly characterize the long-term evolution of the binary and disk. The magnitude of their calculated l_0 is not large enough to cause orbital expansion, in contrast to our results (see Section 6.8.2).

A most striking feature of Fig. 6.19 is that l_0 is, in general, a non-monotonic function of $e_{\rm B}$. For each value of α, the minimum of l_0 occurs at the binary eccentricity ($e_{\rm B} = 0.4$ for $\alpha = 0.1$ and $e_{\rm B} = 0.2$ for $\alpha = 0.05$) for which the inner disk becomes apsidally aligned with the binary (see Section 6.5, particularly Figs. 6.8-6.9). At both lower and higher binary eccentricities, the inner disk precesses coherently, giving rise to a larger value of l_0. We can only speculate that this remarkable inter-dependence of the long-term disk variability and the global disk eigenvalue l_0 is the result of the intricate balance between the different contributions to $\dot{J}$ and the details of the accretion dynamics.

In this paper we have only considered binaries with equal masses or modest mass ratios ($q_{\rm B} \geq 0.2$), for which we have found l_0 is mostly positive (see Table 6.1). Our simuations do not apply to binaries with more extreme mass ratios, for which an annular gap is opened by the (small-mass) secondary instead of a common cavity surrounding both components of the binary (D'Orazio et al. 2016). For such extreme mass ratios, the binary may lose angular momentum to the disk, analogous to Type II planet migration (e.g., Lin & Papaloizou 1986). Another caveat of our study is that we have neglected the self-gravity of the disk, which may play a role in the angular momentum transfer to the binary (e.g., Cuadra et al. 2009). This amounts to assuming that the disk mass $\pi r^2\Sigma$ (inside a few $a_{\rm B}$'s) is much less than the mass of the binary $M_{\rm B}$. Binary black holes

embedded in a massive nuclear disk may suffer angular momentum loss to the disk due to dynamical friction (e.g., Escala et al. 2005).

6.8 Summary and Discussion

6.8.1 Summary of Results

We have carried out long-term, two-dimensional simulations of viscous circumbinary disks, for a range of binary eccentricities, mass ratios, and viscosity parameters. We have focused on the scenario in which mass is supplied to the outer disk at a fixed rate, and examined the quasi-steady state of the disk, as characterized by a steady flow of mass and angular momentum throughout the disk. Our key results are:

- The structure of the inner disk is characterized by the radius $r_{\rm peak} = (3 - 5)a_{\rm B}$ (where $a_{\rm B}$ is the binary semi-major axis), at which the azimuthally-averaged surface density reaches a maximum, and another radius, $r_{0.1} = (1.6 - 2.6)a_{\rm B}$, at which the surface density falls to 10 per cent of its value at $r_{\rm peak}$. While $r_{\rm peak}$ is non-monotonic in $e_{\rm B}$, $r_{0.1}$ strictly increases with $e_{\rm B}$ (see Fig. 6.3), and is a good representation of the "truncation radius" of the disk, as it agrees with the theory of inner disk truncation due to the clearing of a gap at increasingly higher order Lindblad resonances (as a function of $e_{\rm B}$) to within 20 per cent.
- The mass accretion onto the central binary is highly variable (see Fig. 6.5), with the dominant period of variability either equal to the binary period

P_B or about $5P_B$ ("low-frequency variability"). The low-frequency variability occurs for $e_B \lesssim 0.05$ and $q_B \gtrsim 0.2$, and is associated with the Keplerian motion of density "lumps" that are continuously formed and destroyed near the inner edge of the disk (slightly interior to r_{peak}). The accretion variability at P_B occurs for $e_B \gtrsim 0.05$, in which case no lump is formed at the inner disk (see Fig. 6.6).

- The inner region of the disk ($3a_B \lesssim r \lesssim 6a_B$) is generally eccentric. For both low ($e_B \lesssim 0.2$) and high ($e_B \gtrsim 0.4$) binary eccentricities, the eccentric disk coherently precesses around the binary with a period of several hundred binary orbits. For intermediate eccentrities ($e_B = 0.2 - 0.4$, depending on the disk viscosity parameter), the eccentric disk instead becomes apsidally locked with the binary (see Figs. 6.8-6.9). The behavior of the inner eccentric disk may be explained in a linear, one-dimensional framework, using a combination of eccentricity excitation at multiple eccentric Lindblad resonances, the strengths of which depend on e_B, and viscous eccentricity damping (see Section 6.6).

- Although the "instantaneous" mass accretion rate across the disk is highly variable (see Fig. 6.13), the time-averaged accretion rate (averaged over hundreds of P_B, corresponding to the precession period of the inner eccentric disk), is constant across the disk, and approximately equals the mass supply rate at the outer radius of the disk (see Fig. 6.14). Although this result contradicts several previous claims, it does agree with the recent simulations using the moving mesh code AREPO (Muñoz & Lai 2016) that resolve accretion onto individual components of the binary.

- Our most important finding concerns the time-averaged angular momentum transfer rate from the disk to the binary. This angular momentum

transfer includes contributions from mass accretion, viscous stress and gravitational torque between the binary and the disk. We find that, in the quasi-steady state, the specific angular momentum transferred to the binary, $l_0 = \langle \dot{J} \rangle / \langle \dot{M} \rangle$, depends on the binary eccentricity in a non-monotonic manner (Fig. 6.19; see Table 6.1 for a list of the computed l_0 values). Contrary to many previous claims and assumptions, we find that l_0 is positive for most $e_{\rm B}$'s, implying that the binary receives net angular momentum. The minimum l_0 occurs at intermediate binary eccentricities (around $0.2 - 0.4$, depending on the viscosity parameter), and corresponds to the regime where the inner eccentric disk is apsidally aligned with the binary (see Section 6.7.3 for discussion).

6.8.2 Astrophysical Implications

Binary Orbital Evolution

As noted before (Section 6.1 and Section 6.7.3), binary-disk interaction has long been suggested to play an important role in driving the orbital decay of supermassive black hole (SMBH) binaries. The rate of change of the binary angular momentum is related to the rate of change of its orbital elements, $\dot{a}_{\rm B}$ and $\dot{e}_{\rm B}$, as well as the mass accretion rate of the individual stars, $\dot{M}_1$ and $\dot{M}_2$, according to

$$\frac{\dot{J}_{\rm B}}{J_{\rm B}} = \frac{\dot{M}_1}{M_1} + \frac{\dot{M}_2}{M_2} - \frac{1}{2}\frac{\dot{M}_{\rm B}}{M_{\rm B}} + \frac{1}{2}\frac{\dot{a}_{\rm B}}{a_{\rm B}} - \left(\frac{e_{\rm B}}{1-e_{\rm B}^2}\right)\dot{e}_{\rm B}. \tag{6.38}$$

We cannot track the mass accretion rates of the individual stars in the simulations presented in this paper, but for an equal mass binary, we may assume that on average, $\dot{M}_1 = \dot{M}_2 = \dot{M}_{\rm B}/2$, due to symmetry (this is true on average, but the

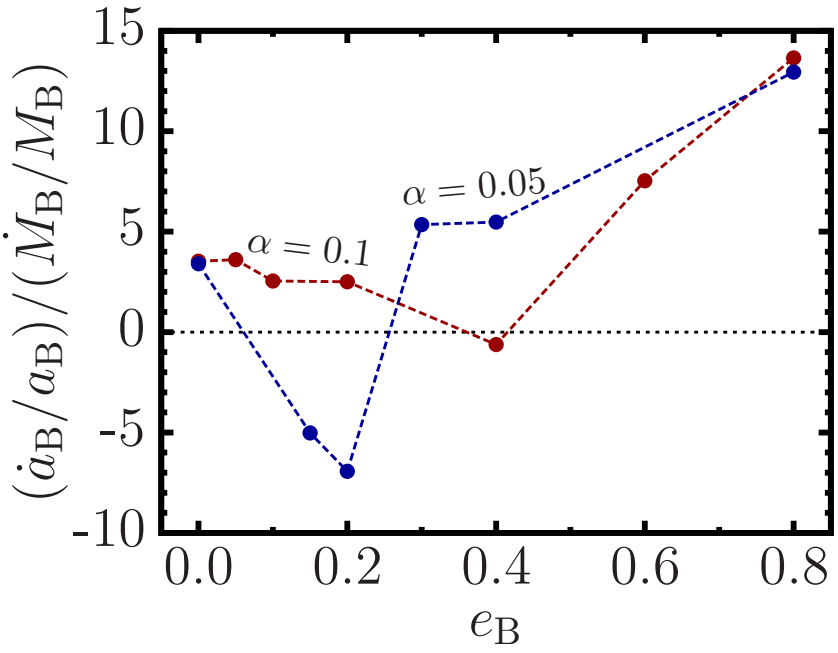

Figure 6.20: The rate at which the binary semi-major axis changes, $\dot{a}_{\rm B}/a_{\rm B}$ (Eq. 6.39), normalized by the rate at which its mass grows, $\dot{M}_{\rm B}/M_{\rm B}$, as inferred from the values of l_0 shown in Fig. 6.19. We have assumed that the members of the equal mass binary take on equal shares of the accreted mass, and that the binary eccentricity does not change. Under these assumptions, the semi-major axis of the binary grows with time for a wide range of $e_{\rm B}$.

symmetry can be temporarily broken in an alternating fashion due to precession, e.g., Dunhill et al. 2015; Muñoz & Lai 2016). Solving the full evolution of the binary orbit also requires knowledge of the rate at which it gains or loses energy, the determination of which is beyond the scope of this work. If instead we simply assume that $\dot{e}_{\rm B} = 0$, the orbital evolution is described by

$$\frac{\dot{a}_{\rm B}}{a_{\rm B}} = 8\left(\frac{l_0}{l_{\rm B}} - \frac{3}{8}\right)\frac{\dot{M}_{\rm B}}{M_{\rm B}}, \tag{6.39}$$

where $l_{\rm B} = [GM_{\rm tot}a_{\rm B}(1-e_{\rm B}^2)]^{1/2}$ is the specific angular momentum of the binary. Under this assumption, the sign of $\dot{a}_{\rm B}/a_{\rm B}$ is determined by whether l_0 is larger or smaller than $3l_{\rm B}/8$. Figure 6.20 shows the orbital evolution rate for equal-mass binaries, using the values of l_0 determined in our numerical simulations (see Table 6.1 and Fig. 6.19). Typically, $\dot{a}_{\rm B}$ is positive, meaning that the binary semi-major axis grows with time. We find two cases for which the binary shrinks, both corresponding to the minimum of $l_0(e_{\rm B})$, which is also associated with apsidal alignment of the eccentric disk and binary. We reiterate that Eq. 6.39 applies only when $\dot{e}_{\rm B} = 0$. Eccentricity damping ($\dot{e}_{\rm B} < 0$) would reduce $\dot{a}_{\rm B}$.

This result is contrary to the common assumption that interactions with a surrounding disk lead to binary shrinkage (e.g., Begelman et al. 1980; Armitage & Natarajan 2002; MacFadyen & Milosavljević 2008). Therefore, at least for binary black holes with comparable masses, circumbinary disks may not provide a solution to the "final parsec problem" for SMBH binaries. As a caveat, we reiterate that the simulations in this paper were restricted to binaries with equal masses or modest mass ratios, and we have also assumed that the inner disk (within $\sim 10a_{\rm B}$) has a negligible mass compared to the binary (see Section 6.7.3). More systematic studies of binaries with more extreme mass ratios and massive disks are warranted.

Circumbinary Planet Formation

Circumbinary protoplanetary disks, a number of which have been observed (e.g., Dutrey et al. 1994; Mathieu et al. 1997; Simon et al. 2000), are the birthplaces of circumbinary planets, such as those discovered by the *Kepler* mission (e.g., Doyle et al. 2011; Welsh et al. 2012; Orosz et al. 2012). Many of these planets have semi-major axes a few times larger than that of the binary, and are therefore close to the dynamical stability limit (Holman & Wiegert 1999). This radius is also close to the inner truncation radius of the disk, which may play a role in halting planetary migration (e.g., Nelson 2003; Pierens & Nelson 2007; Kley & Haghighipour 2014, 2015).

The disk truncation radius increases with binary eccentricity (see Fig. 6.3). The gradients of the surface density profile in the inner disk are also sensitive to the binary eccentricity, becoming steepest when it takes on intermediate values. These steep density profiles may affect planet migration in these regions.

The growth of planetesimals may be inefficient close to the binary as a result of its gravitational perturbations (e.g., Moriwaki & Nakagawa 2004; Paardekooper et al. 2012; Meschiari 2012; Rafikov 2013; Silsbee & Rafikov 2015). The eccentricity of planetesimals embedded in the disk is damped due to gas drag, but it cannot reach zero, since the finite eccentricity of the disk itself serves as a lower limit. The average relative velocity between planetesimals is approximately $v_{\rm rel} \approx e v_{\rm K}$, where $v_{\rm K}$ is the Keplerian orbital velocity. At a separation of $(3-4)a_{\rm B}$ from the binary, near the dynamical stability limit where many of the Kepler circumbinary planets reside, we find that the disk has an eccentricity of $0.05-0.2$, which for a binary separation of 0.2AU, corresponds to relative planetesimal velocities of several km/s. As this is much larger than their escape

velocities (e.g., a few m/s for 10km bodies), collisional growth of planetesimals is likely to be strongly inhibited in this region. However, we find that for intermediate binary eccentricities, the eccentric disk becomes apsidally aligned with the binary. If, in these cases, planetesimal orbits closely follow those of fluid elements in the disk (due to the effects of gas drag), then the planetesimals may be apsidally aligned with one another. As a result, they may have much smaller relative velocities, allowing collisional growth to occur. A full treatment of both the gas and planetesimal dynamics is required to assess the plausibility of this outcome.

6.8.3 Limitations and Prospects

Throughout this work, we have studied the accretion flow around binaries while discarding the region of the computational domain where individual circumsingle disks would form around each component of the binary. This simplification makes the problem tractable in cylindrical-coordinate grids. In addition, ignoring the inner region makes the long-term integrations presented here feasible, as it directly imposes a lower limit on the shortest computational time-step. However, as a consequence, the details of accretion onto the individual members of the binary – such as mass and momentum transfer – are lost. Simulating this co-orbital region is key to understanding the mass-growth and orbital evolution of the central binary, but it represents a significant computational challenge. Long-term simulation of individual accretion in binaries while retaining the features of FARGO-like finite-volume codes such as PLUTO has only been accomplished recently (Farris 2014, Muñoz & Lai 2016). These simulations are now possible thanks to the implementation of novel meshing algorithms

in the ground-breaking codes DISCO (Duffell & MacFadyen 2013, Duffell 2016) and AREPO (Springel 2010, Muñoz et al. 2014). In future work (Muñoz, Miranda & Lai, in prep), we will complement the results of our PLUTO simulations with AREPO simulations. As AREPO and PLUTO are both finite-volume schemes (albeit with different-order reconstruction), the cell-centred primitive variables of a PLUTO snapshot can be directly mapped onto an AREPO unstructured grid simulation and then evolved onwards. With virtually identical initial conditions, the mass and angular momentum exchange between the disk and the binary can be contrasted for these two numerical approaches, thus providing a test and validation for the conclusions presented in this work.

In particular, it is of interest to test the robustness of l_0. Provided that the simulation is in a quasi-steady state, the value of l_0 is largely determined by the surface density profile, Σ [Eq. (6.37), with both $F(r)$ and $\nu(r)$ being the same for different numerical methods, within the errors of the hydrodynamical solver]. Since the diode boundary condition at $r = r_{\rm in}$ in our PLUTO simulations cannot capture the "sloshing" nature of the gas dynamics around eccentric binaries at that location, it is possible that even though $\dot{M}(r) = \text{constant}$ in both sets of simulations, a slightly different value of Σ at $r \approx r_{\rm in}$ might change the value of l_0. However, the qualitative agreement in $\dot{M}(t)$ between the two sets of simulations (see Figs. 2 and 5 of Muñoz & Lai 2016) for different eccentricities and locations within the disk (except at $r = r_{\rm in}$) is encouraging. The fact that the value of l_0 calculated in our PLUTO simulations is independent of $r_{\rm in}$ (see Fig. 6.17) also suggests the robustness of our l_0 values.

CHAPTER 7

TRAPPING OF LOW-MASS PLANETS NEAR THE INNER EDGES OF PROTOPLANETARY DISKS

7.1 Introduction

Tidal interactions of planets with their natal disks lead to angular momentum exchange and orbital migration (Goldreich & Tremaine 1979, 1980; Lin & Papaloizou 1979; Ward 1986, 1997). The direction of the orbital migration is usually inwards, although this depends on the structure, thermodynamics, and radiative and turbulent properties of the disk (see Kley & Nelson 2012; Baruteau et al. 2014). For low-mass planets, with insufficient mass to modify the surface density profile of the disk, the resulting "Type I" migration is rapid, and proceeds until the planet reaches the inner edge of the disk or the magnetospheric cavity.

In Type I migration theory, the disk responds strongly to the gravitational potential of the planet at Lindblad resonances (LRs), which are locations at which the natural radial epicyclic frequency of the disk is commensurate with the pattern frequency of the planet. Waves excited at either outer Lindblad resonances (OLRs), which carry positive angular momentum, or at inner Lindblad resonances (ILRs), which carry negative angular momentum, propagate away from the planet until they are damped either viscously (e.g., Takeuchi et al. 1996) or nonlinearly (e.g., Goodman & Rafikov 2001). Migration is a result of the dif-

This chapter is adapted from Miranda & Lai (in prep).

ferential Lindblad torque, i.e., the asymmetry between the negative torque applied to the planet by OLRs and the positive torque by ILRs. In standard theory, the disk is considered to be infinite in extent, so that waves travel exclusively away from the planet and damp far from it. In numerical studies of planet-disk interaction (e.g., de Val-Borro et al. 2006), this is achieved by imposing wave damping zones at the edges of the computational domain, so that all waves are outgoing, and wave reflection from the edges of the domain is prohibited.

Tsang (2011; hereafter T11) demonstrated that, if the inner edge of the disk is partially reflective, waves launched by the planet at ILRs may be resonant with the inner edge of the disk, enhancing their positive torque on the planet. If the resonant gain of these torques is sufficiently large, then the sum of ILR torques may be equal in magnitude to the OLR torques, resulting in a vanishing differential Lindblad torque. This leads to the existence of migration traps in the disk, at which the migration of the planet is halted, with a semi-major axis several times larger than the radius of the inner edge of the disk. This mechanism may be somewhat modified by the corotation torque, which arises due to interaction of gas in the co-orbital region of the planet (provided that it is not saturated, see Kley & Nelson 2012).

Protoplanetary disks may either extend to the stellar surface, or be truncated by the stellar magnetic field. In the latter case, accretion onto the central star is complex and three-dimensional inside of the truncation radius, involving the funneling of disk material along the stellar magnetic field lines. In either case, the inner edge of the disk (i.e., disk-star or disk-magnetosphere interface) may present a semi-rigid "wall" to incoming waves (e.g., Fu & Lai 2012), allowing them to be reflected, and making the migration halting mechanism described

by T11 possible.

In this paper, we numerically investigate the migration of a low-mass planet ($\lesssim 10M_\oplus$) near the inner edge of a disk, using two-dimensional, viscous hydrodynamics simulations. We assume that the disk is truncated at the stellar corotation radius, in a configuration such that no angular momentum is transfered to the star. We show that Type I migration is halted at several times the radius of the inner edge of the disk.

7.2 Problem Setup

We consider a planet with mass M_p and semi-major axis a_p embedded in a gaseous disk orbiting a star of mass M_*. The planet-to-star mass ratio is is $q = M_\mathrm{p}/M_*$. The disk is described in (cylindrical) polar coordinates (r, ϕ) by the surface density Σ and velocity (u_r, u_ϕ). The disk is geometrically thin, and rotates at approximately the local Keplerian rate $\Omega_\mathrm{K} = (GM_*/r^3)^{1/2}$. Its radial extent is r_in to $r_\mathrm{out} = 7r_\mathrm{in}$, and the orbital period at r_in, $P_\mathrm{in} = 2\pi/\Omega_\mathrm{K}(r_\mathrm{in})$, is taken as the reference unit of time. The equation of state is locally isothermal, with $p = \Sigma c_\mathrm{s}^2(r)$, where p is the height-integrated pressure, and

$$c_\mathrm{s}(r) = hr\Omega_\mathrm{K} \tag{7.1}$$

is the sound speed. The disk aspect ratio, $h = H/r$, is chosen to have the constant value $h = 0.05$, unless otherwise specified. The kinematic viscosity is described by the α prescription,

$$\nu = \alpha c_\mathrm{s} H, \tag{7.2}$$

and we choose several different values of α throughout this study.

7.2.1 Disk Structure

The disk is initialized in an equilibrium state, which is not strongly modified by the presence of the (low-mass) planet (i.e., the disk and planet are in the Type I migration regime). We adopt a state with a constant mass accretion rate, $\dot{M} = -2\pi r \Sigma u_r$. The disk structure is determined by conservation of angular momentum, which takes the form

$$\dot{M} l + 2\pi r^3 \nu \Sigma \frac{\mathrm{d}\Omega}{\mathrm{d}r} = \dot{M} l_0, \tag{7.3}$$

where the first term on the left-hand side represents inward advection of angular momentum, and the second term represents outward transport of angular momentum due to viscous stress. The quantity l_0 on the right-hand side is an eigenvalue, whose value is equal to the (constant) ratio of the angular momentum accretion rate and mass accretion rate through the disk (and onto the central star), $\dot{J}/\dot{M}$ (e.g., Popham & Narayan 1991). The profiles of surface density and radial velocity satisfying this equation are

$$\Sigma(r) = \frac{\dot{M}}{3\pi\nu}\left(1 - \frac{l_0}{l}\right), \tag{7.4}$$

$$u_r(r) = -\frac{3}{2}\frac{\nu}{r}\left(1 - \frac{l_0}{l}\right)^{-1}. \tag{7.5}$$

Additionally, the angular velocity is approximately Keplerian, but modified by the pressure gradient,

$$u_\phi(r) = \left(v_\mathrm{K}^2 + \frac{r}{\Sigma}\frac{\mathrm{d}p}{\mathrm{d}r}\right)^{1/2}, \tag{7.6}$$

where $v_\mathrm{K} = r\Omega_\mathrm{K}$ is the Keplerian orbital velocity.

Inner Boundary

The value of the eigenvalue l_0 is determined by the details of the inner edge of the disk, which are specified by the numerical inner boundary conditions. For non-magnetic stars, the disk stops at the stellar surface, $r_{\rm in} = R_*$. More likely, for magnetic stars, the disk is truncated at the magnetosphere radius, at which the stress due to the (dipolar) stellar magnetic field dominates over the accretion stress,

$$r_{\rm m} = k\left(\frac{B_*^4 R_*^{12}}{GM_*\dot{M}^2}\right)^{1/7}, \tag{7.7}$$

where B_* is the stellar surface magnetic field strength and $k \approx 1$ is a constant. In both cases, to avoid complications associated with boundary layers, we assume that the flow inside inside $r_{\rm in}$ rotates at $\Omega_{\rm K}(r_{\rm in})$, i.e., $r_{\rm in}$ coincides with the corotation radius of the star,

$$r_{\rm co} = \left(\frac{GM_*}{\Omega_*^2}\right)^{1/3}. \tag{7.8}$$

The plausibility of such a configuration is supported by analytic and numerical studies of the interactions of disks with magnetic stars (Ghosh & Lamb 1979; Koenigl 1991; Ostriker & Shu 1995; Long et al. 2005). In this state, there is no angular momentum transfered to the star, which fixes the value of the eigenvalue, $l_0 = 0$. Equations 7.4 and 7.5 then become

$$\Sigma(r) = \Sigma_0\left(\frac{r}{r_{\rm in}}\right)^{-1/2}, \tag{7.9}$$

where $\Sigma_0 = \dot{M}/[3\pi\nu(r_{\rm in})]$, and

$$u_r(r) = -\frac{3}{2}\alpha h^2 v_{\rm K}. \tag{7.10}$$

In general, fluid perturbations around $r_{\rm in}$ are complicated, and depend on the dynamics of the flow inside $r_{\rm in}$ (see Tsang & Lai 2009; Fu & Lai 2012). For simplicity, we assume that at $r_{\rm in}$, the orbital velocity is fixed at its Keplerian value,

while the surface density and radial velocity are fixed are at their equilibrium values (i.e., equations 7.9 and 7.10 evaluated at $r_{\rm in}$):

$$\left(\Sigma, u_r, u_\phi\right)_{r_{\rm in}} = \left[\Sigma_0, \left(-\frac{3\nu}{2r}\right)_{r_{\rm in}}, v_{\rm K}(r_{\rm in})\right]. \tag{7.11}$$

Disk Mass and Planet Migration

The choice of the value of Σ_0 determines the migration rate of the planet. The total (Lindblad + corotation) torque on the planet can be written as $\Gamma_{\rm tot} = C\Gamma_0$, where

$$\Gamma_0 = \left(\frac{M_{\rm p}}{M_*}\right)^2 \left(\frac{H}{a_{\rm p}}\right)^{-2} \Sigma_{\rm p} a_{\rm p}^4 \Omega_{\rm p}^2 \tag{7.12}$$

is the characteristic Type I migration torque [here $\Sigma_{\rm p} = \Sigma(a_{\rm p})$], and the order unity coefficient C depends on the structure of the disk. For an infinite, two dimensional disk with $\Sigma \propto r^{-1/2}$ and $c_{\rm s} \propto r^{-1/2}$, $C = 1.47$ (D'Angelo & Lubow 2010). The orbital evolution of the planet is determined by setting $\Gamma_{\rm tot} = \dot{L}_{\rm p}$, where $L_{\rm p}$ is the angular momentum of the planet. Thus the semi-major axis decays exponentially (this will be modified due to the partially reflective inner boundary), with a timescale given by

$$\begin{aligned} t_{\rm mig} &= \frac{h^2}{4\pi C q}\left(\frac{\Sigma_0 r_{\rm in}^2}{M_*}\right)^{-1} P_{\rm in} \\ &= 1.35\times 10^4 \left(\frac{q}{10^{-5}}\right)^{-1}\left(\frac{h}{0.05}\right)^2\left(\frac{\Sigma_0 r_{\rm in}^2}{10^{-3}M_*}\right)^{-1} P_{\rm in}. \end{aligned} \tag{7.13}$$

We choose $\Sigma_0 = 10^{-3} M_*/r_{\rm in}^2$ for the surface density scaling factor, and have scaled equation 7.13 accordingly. Note that this value of Σ_0 is about 50 times larger than the one corresponding to a minimum-mass solar nebula profile (Hayashi 1981), $6.0\times 10^{-5} M_*/r_{\rm in}^2$ (taking $r_{\rm in} = 0.1$ AU), which would lead to a longer migration timescale ($2.26\times 10^5 P_{\rm in}$) for our computation. However, it is

only about two times larger than the value corresponding to the "minimum-mass extrasolar nebula" of Chiang & Laughlin (2013).

For most of our simulations, the initial semi-major axis of the planet is $a_0 = a_p(t = 0) = 3r_{in}$. Based on the results of T11, this is an approximate upper limit for the semi-major axis at which migration may be halted. However, we find that the migration halting mechanism is somewhat more robust than indicated by the analytical results, and we find several cases for which the planet migrates negligibly when initialized at $3r_{in}$. In these cases, the simulation is repeated with a larger a_0, to check whether or not there is a larger semi-major axis at which the migration of the planet can be halted. Indeed, for our standard disk thickness, halting slightly beyond $3r_{in}$ is possible in some cases, while for thicker disks, halting significantly beyond $3r_{in}$ is sometimes possible, requiring a much larger a_0.

7.2.2 Wave Damping Zones

Near the outer disk boundary ($r/r_{in} = 6.4 - 7.0$), we apply the wave damping conditions (e.g., de Val-Borro et al. 2006), in which the fluid variables are damped to their initial conditions on the local orbital timescale. This prevents wave reflection at the outer boundary, so that only outgoing waves may propagate (physically, this represents the existence of an extended disk beyond the outer boundary). At the inner boundary, we do not apply the damping conditions, as we are explicitly interested in reflection of waves by the inner edge. However, for comparison, we have also performed several runs in which wave damping near the inner boundary ($r/r_{in} = 1.0 - 1.3$) is included, to mimic the

standard numerical planet-disk interaction setup, in which the disk extends to much smaller radii.

7.2.3 Numerical Method

The two-dimensional viscous hydrodynamic equations of the disk and the dynamics of the planet are solved using the code FARGO3D (Benítez-Llambay & Masset 2016). The gravitational potential of the planet is softened over a length of $\epsilon = 0.6H$. The shift of resonances due to the fact that the planet feels the gravity of the disk, while the disk does not, is corrected using the method described in Baruteau & Masset (2008). Unless otherwise stated, the dimensions of the numerical grid are $N_r \times N_\phi = 477 \times 1536$ (see Section 7.3.5 for a resolution study), with logarithmic spacing in r and uniform spacing in ϕ, so that the grid cells are approximately square ($\Delta r \approx r\Delta\phi$) everywhere. The planet is smoothly introduced to the disk by ramping its mass up from zero over the first $50P_{\rm in}$ (approximately 10 orbits of the planet at its initial $a_{\rm p} = 3r_{\rm in}$).

7.3 Results

7.3.1 Migration and Trapping

The migration (semi-major axis as a function of time) of the planet for several different mass ratios in disks with different values of the viscous α parameter is shown in Figure 7.1. The planet initially migrates inwards at a rate approximately equal to the migration rate in an infinite disk (equation 7.13), on aver-

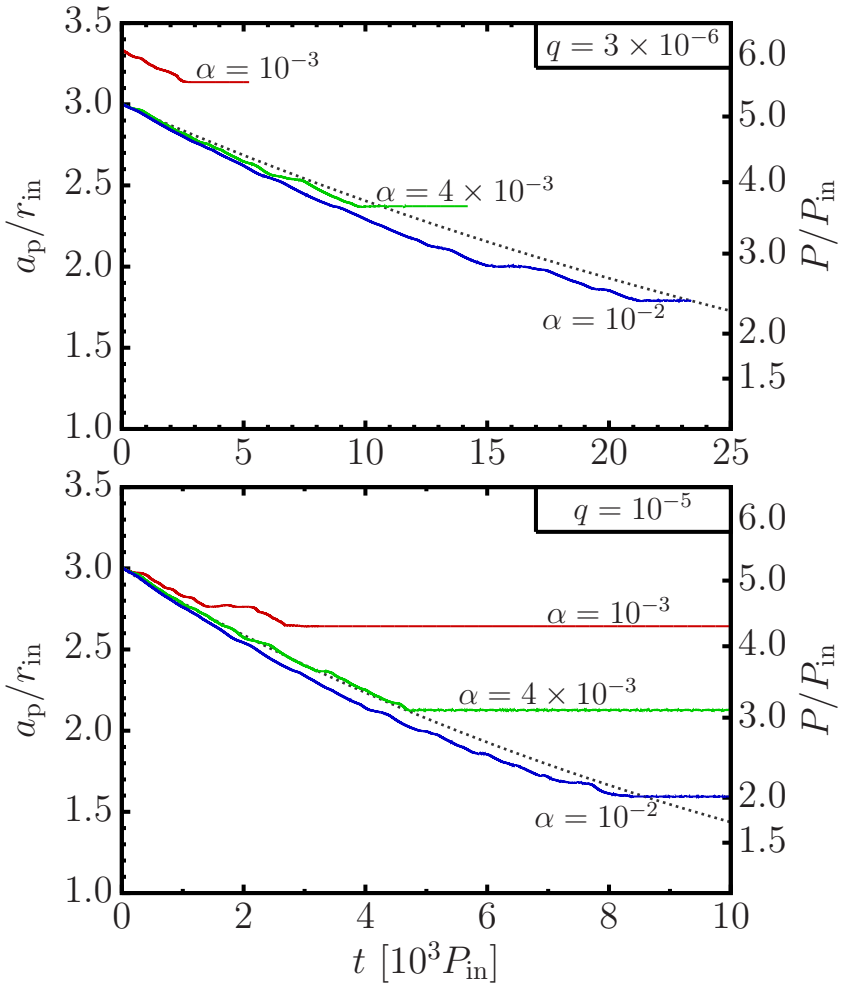

Figure 7.1: The semi-major axis of the planet as function of time for $q = M_\mathrm{p}/M_* = 3 \times 10^{-6}$ (top panel) and $q = 10^{-5}$ (bottom panel), with different values of the viscosity parameter, α (solid lines). The dotted lines correspond to the migration of the planet in an infinite disk. Note that the scale of the x-axis differs between the two panels, and that in the top panel, for the case with $\alpha = 10^{-3}$, the planet is initialized with a slightly larger semi-major axis than in the other cases.

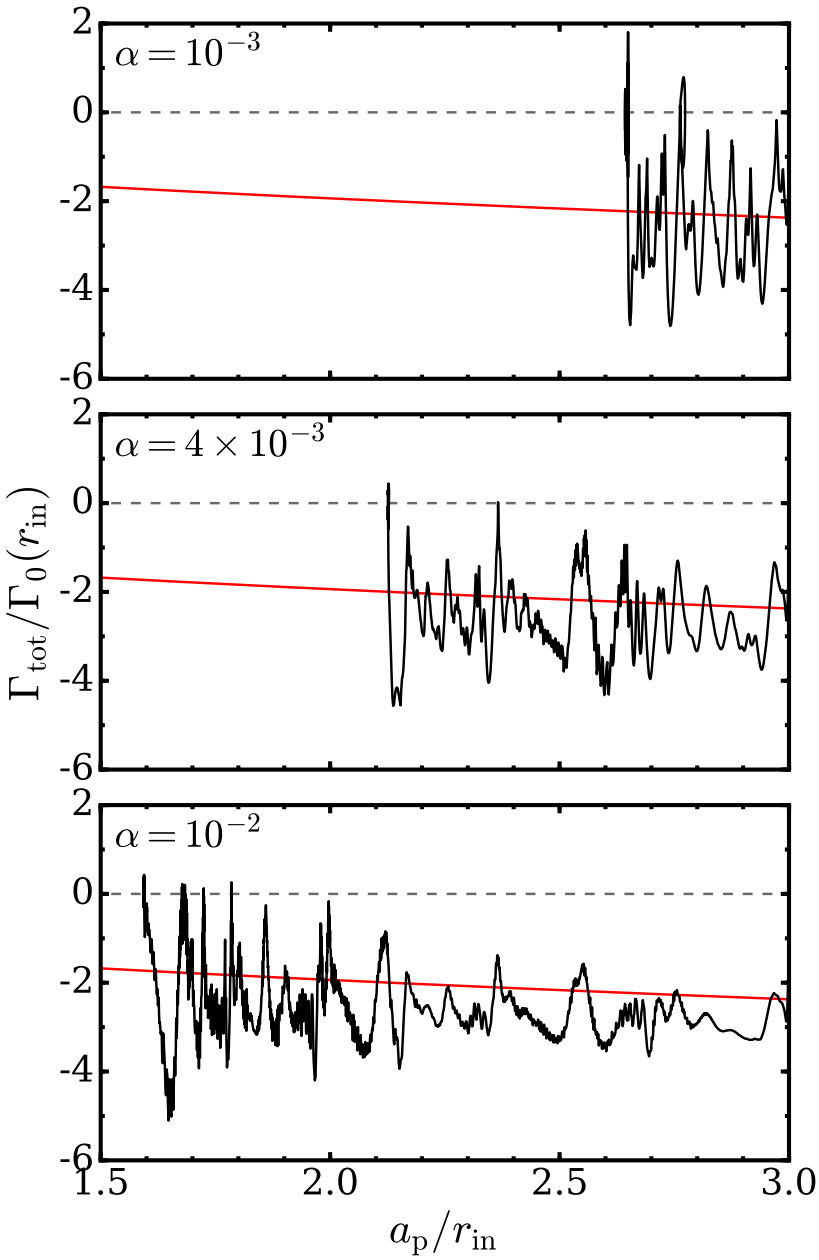

Figure 7.2: Total torque (black curves) on the planet for the case with $q = 10^{-5}$, as a function of semi-major axis, for three different values of α (different panels). The planet moves from right to left in these plots. The torque has been time-averaged over intervals of $5P_{\rm in}$ in order to filter out short-timescale variations. The red line represents the torque on a planet undergoing normal Type I migration. The location in the disk at which the planet ultimately stops migrating is coincident with a zero-crossing of $\Gamma_{\rm tot}$.

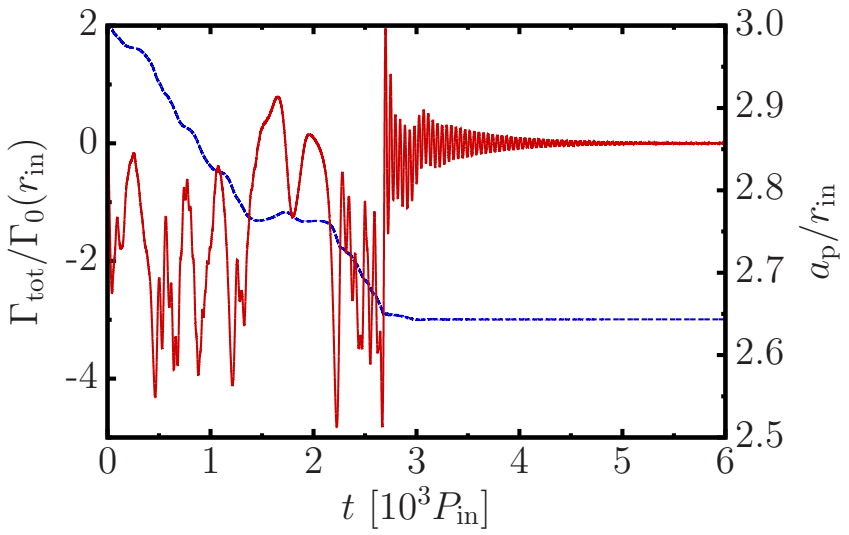

Figure 7.3: Total torque on the planet (solid line, left axis) and semi-major axis (dashed line, right axis) as a function of time for the case with $q = 10^{-5}$ and $\alpha = 10^{-3}$.

age. However, the instantaneous migration rate varies relative to the average on timescales of hundreds to thousands times $P_{\rm in}$. Eventually, the planet encounters a "trap" and stops migrating, with the semi-major axis remaining constant for the remainder of the simulation (at least several thousand $P_{\rm in}$ for all cases).

The total torque on the planet as a function of semi-major axis is shown in Figure 7.2, for the cases with $q = 10^{-5}$ and several values of α (cf. Figure 3 of T11). The standard Type I migration torque is also shown for comparison. The torque is an oscillatory function of $a_{\rm p}$; averaged over a range of $a_{\rm p}$, the torque is equal to the the standard torque, but it can also be significantly larger or smaller (in magnitude) at different locations in the disk. In particular, as described in T11, the torque has zero-crossings, corresponding to locations at which migration is halted. See Figure 7.3 for the torque as a function of time, which decays towards zero as the planet settles into the trap. The calculations presented in T11 suggest that the spatial variations in torque and presence of zero-crossings are largely

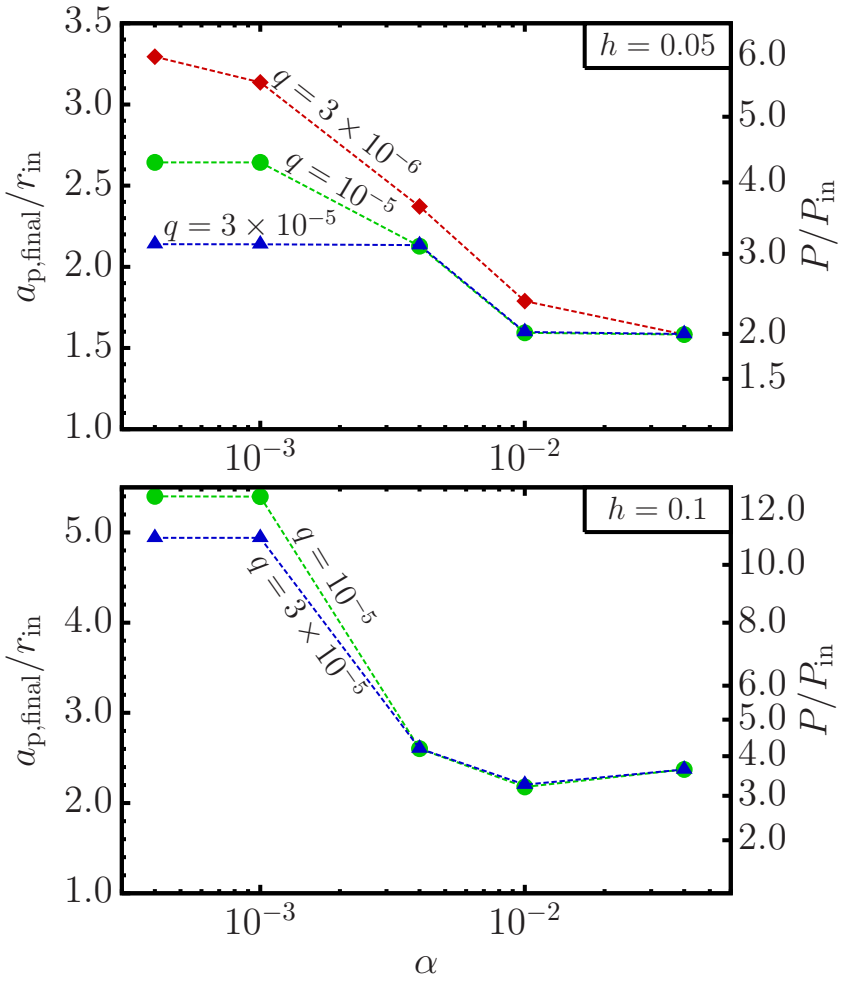

Figure 7.4: The final semi-major axis of the planet, $a_{\rm p,final}$, as a function of α, for different values of $q = M_{\rm p}/M_*$. In the top panel, the disk aspect ratio is $h = 0.05$ (the same value used in all other figures in this paper), while in the bottom panel, we adopt the larger value $h = 0.1$.

suppressed for $\alpha = 10^{-3}$ (for $h = 0.05$, as in our simulations). However, we find that these features are present for values of α as large as 0.04.

In some cases, there are other zero crossings which the planet migrates through without stopping. These correspond to intervals of time, for example $t \approx (1500 - 2500)P_{\rm in}$ for the case with $q = 10^{-5}$ and $\alpha = 10^{-3}$, for which migration is dramatically slowed down, but not completely stopped. At these points, the planet spends an extended amount of time with a nearly constant semi-major axis, but eventually breaks out of the "temporary" migration trap and continues to migrate inwards, until ultimately encountering a "permanent" trap. This behavior can be attributed to short-term variations in the torque on the orbital period and grid cell crossing time of the planet (the latter is a numerical effect, see Section 7.3.5). If the amplitude of the torque associated with these variations is larger than the slowly-varying torque associated with a migration trap, then the planet can in principle escape from a trap.

The final semi-major axis of the planet, $a_{\rm p,final}$, is shown in Figure 7.4 (top panel), for planets with $q = 3 \times 10^{-6}$, 10^{-5}, and 3×10^{-5} (i.e., $M_\oplus$, $3M_\oplus$, and $10M_\oplus$, if $M_* = 1M_\odot$), with a viscous α in the range $4 \times 10^{-4} - 4 \times 10^{-2}$ (note that several of these cases are not shown in Figure 7.1). Generally, $a_{\rm p,final}$ decreases as either q or α increases. For each q, there is a maximum $a_{\rm p,final}$, whose value decreases with q, which is achieved when α is sufficiently small ($\lesssim 10^{-3}$). As α is increased from 10^{-3} to 10^{-2}, $a_{\rm p,final}$ decreases. For $\alpha \gtrsim 10^{-2}$, $a_{\rm p,final}$ is approximately constant, at a value of about $1.59r_{\rm in}$. Although this is very close to the location of the 2:1 orbital commensurability with the inner edge of the disk, it has no physical significance and is merely a numerical coincidence, since we find that $a_{\rm p,final}$ can be shifted away from this value with small changes in the value of h.

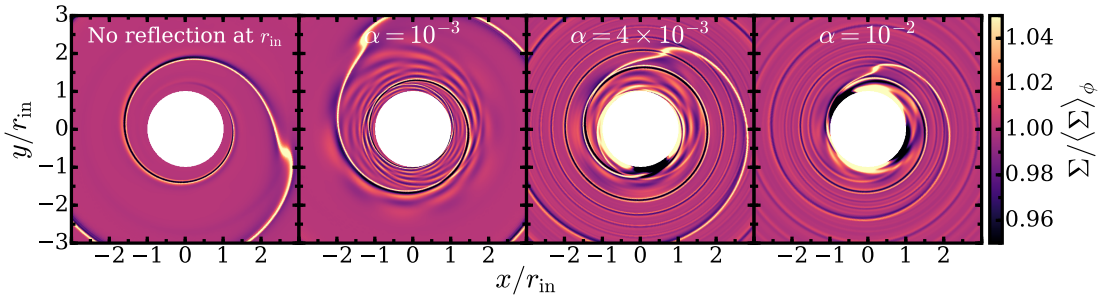

Figure 7.5: Snapshots of the disk surface density, relative to the azimuthally-averaged surface density, for a planet with $q = 10^{-5}$. In the leftmost panel, the planet is undergoing normal Type I migration (i.e., an inner wave damping zone is included). In the other panels, the planet is self-consistently trapped at its final semi-major axis, for different values of α.

Also shown in Figure 7.4 (bottom panel) is $a_{\rm p,final}$ as a function of α for two different planet masses, for the case with a larger disk aspect ratio, $h = 0.1$. The results are qualitatively similar to the case with $h = 0.05$; $a_{\rm p,final}$ decreases with α before reaching a nearly constant minimum value for $\alpha \gtrsim 10^{-2}$ (there is a slight increase in $a_{\rm p,final}$ when α is further increased from this value, although its magnitude is small relative to the total range of $a_{\rm p,final}$). However, $a_{\rm p,final}$ is systematically larger compared to the case with smaller h, and is never smaller than $2.2r_{\rm in}$. For $\alpha \lesssim 10^{-3}$, $a_{\rm p,final}$ can be as large as $5.4r_{\rm in}$ (note that for theses cases, the planet was initialized with a semi-major axis of $5.5r_{\rm in}$, and the outer boundary of the numerical grid was extended to $12r_{\rm in}$).

7.3.2 Wave Reflection

Figure 7.5 shows snapshots of the disk surface density (relative to the local azimuthally-averaged surface density), when the planet is at its final semi-major axis, for the case with $q = 10^{-5}$ and several different values of α. Also shown for

comparison is the case of a freely migrating planet, with an inner wave damping zone included so that wave reflection at the inner boundary is prohibited. For the cases in which the planet is trapped, there are visible standing waves between the inner edge and the orbit of the planet. For $\alpha = 10^{-3}$, the dominant azimuthal number of the reflected waves, $m \approx 8$, can be discerned visually. These reflected waves do not propagate outward beyond the location of the planet. For the larger viscosity parameters, $\alpha = 4 \times 10^{-3}$ and 10^{-2}, in which the planet is trapped closer to the inner edge, the reflected waves have a larger amplitude, and can propagate beyond the orbit of the planet.

7.3.3 Enhancement of Inner Lindblad Torques

We decompose the torque on the planet into components which have different azimuthal numbers and which originate from different parts of the disk, according to

$$\Gamma_m = \pi \int r \Sigma_m(r) \left(\frac{\partial \Phi}{\partial \phi} \right)_m (r) \mathrm{d}r, \tag{7.14}$$

where Σ_m and $(\partial\Phi/\partial\phi)_m$ are the m-components (i.e., parts proportional to $e^{\mathrm{i}m\phi}$) of the surface density and (softened) tidal potential. The integral is taken over either $[r_{\mathrm{in}}, a_{\mathrm{p}} - r_{\mathrm{H}}]$ for torques due to ILRs, $[a_{\mathrm{p}} - r_{\mathrm{H}}, a_{\mathrm{p}} + r_{\mathrm{H}}]$ for corotation resonances (CRs), or $[a_{\mathrm{p}} + r_{\mathrm{H}}, r_{\mathrm{out}}]$ for OLRs. Here $r_{\mathrm{H}} = (q/3)^{1/3} a_{\mathrm{p}}$ is the Hill radius of the planet. Note that these are convenient operational definitions, rather than exact definitions, for the torques due to the different types of resonances. The decomposition is shown in Figure 7.6, for the cases with $q = 10^{-5}$ and two different values of α. Both the case of a (self-consistently) trapped planet, and a planet fixed at the same location, with an inner wave damping zone included (so that wave reflection is prohibited), are shown. The torques due to OLRs and CRs are

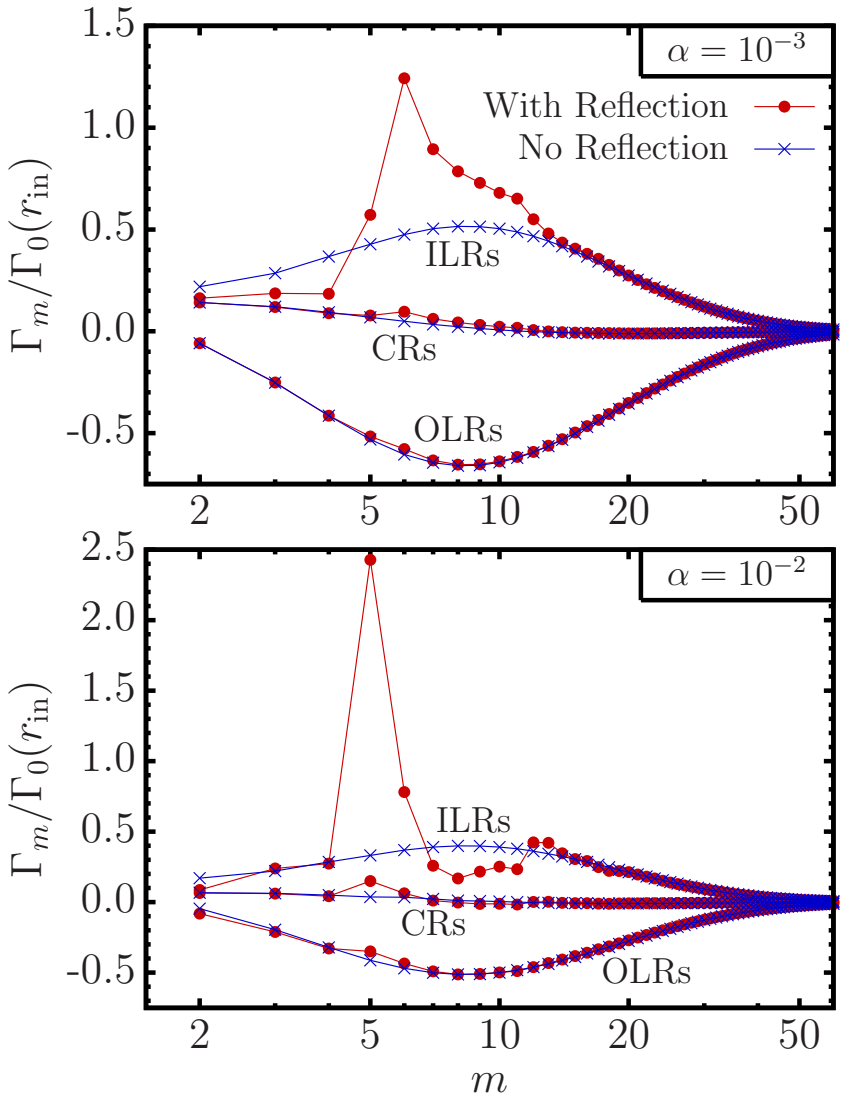

Figure 7.6: The torque on the planet decomposed into components with different azimuthal numbers, Γ_m, and further separated into the contributions from corotation resonances ("CRs") and inner/outer Lindblad resonances ("ILRs/OLRs"), which are operationally defined as originating from within the radial interval $a_{\rm p} \pm r_{\rm H}$, or interior/exterior to this region, respectively. Two cases are shown: the case in which wave reflection at the inner edge is allowed and the planet is self-consistently held in place by a resonant trap (filled circles), and the case in which wave reflection is prohibited by imposing an inner damping condition, and the semi-major axis of the planet is artificially held fixed (crosses). Here $q = 10^{-5}$, and the top and bottom panels correspond to two different values of α. For the case with reflection, the amplitude of the OLRs and CRs is largely unchanged compared to the case without reflection, but the amplitude of the ILRs is enhanced for a range of values of m, specifically $m = 5 - 12$ for $\alpha = 10^{-3}$ and $m = 5 - 6$ for $\alpha = 10^{-2}$.

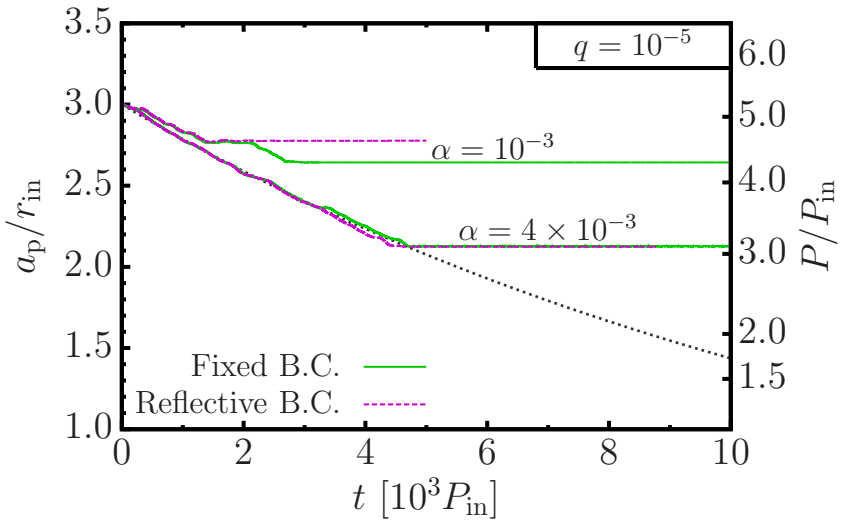

Figure 7.7: The semi-major axis as a function of time for $q = 10^{-5}$, and with different values of α, using two different boundary conditions: the standard "fixed" boundary condition, and the reflective boundary condition (see Section 7.3.4). Neither the details of the migration nor the final semi-major axis of the planet are strongly affected by the particular choice of the inner boundary condition.

mostly unaffected by reflection, while the ILRs are enhanced for a range of m values. For $\alpha = 10^{-3}$, the enhanced ILRs have $m = 5 - 12$, and for $\alpha = 10^{-2}$, they have $m = 5 - 6$. Torque components with small values of m are preferentially enhanced when α is large, since the ILRs with large values of m are located too far from the inner edge for waves launched there to be reflected before they are damped.

7.3.4 Alternative Inner Boundary Condition

The inner boundary conditions we have applied, in which all fluid variables are held at fixed values, represent a simple model of the disk-magnetosphere boundary, which we do not attempt to model in any explicit detail. To assess

the robustness of the planet trapping phenomenon found above, we have investigated an alternative boundary condition, namely a "reflective" boundary condition, in which Σ and u_ϕ are symmetrized across the boundary and u_r is antisymmetrized, so that

$$\left(\frac{\partial \Sigma}{\partial r}, u_r, \frac{\partial u_\phi}{\partial r},\right)_{r_{\rm in}} = (0, 0, 0). \tag{7.15}$$

The results of this numerical experiment, for the cases with $q = 10^{-5}$ and two different values of α, are shown in Figure 7.7. The details of the migration history of the planet are similar, but not identical to the case with the fixed boundary conditions, and ultimately, the migration halts at a nearly the same location. For $\alpha = 10^{-3}$, migration halts at $2.78 r_{\rm in}$, which is the location of a "temporary trap" at which the planet temporarily stalls in our fixed boundary condition simulation. This represents a 5% difference in the final semi-major axis between the two different boundary conditions. For $\alpha = 4 \times 10^{-3}$, the planet stops at $2.13 r_{\rm in}$ regardless of the inner boundary condition. We conclude that our results are not sensitive to the details of the inner boundary condition.

7.3.5 Numerical Resolution

There is a resolution-dependent numerical effect which may influence the migration and trapping of the planet. The torque on the planet varies as it migrates through the numerical grid, with a period equal to the radial grid cell crossing time, $|\Delta r / \dot{a}_{\rm p}|$, where Δr is the grid spacing (Masset & Papaloizou 2003), and with an amplitude which decreases as the resolution increases. To test how numerical resolution affects the final semi-major axis of the planet, we have carried out a resolution study for the case with $q = 10^{-5}$ and $\alpha = 10^{-3}$. Our standard

numerical resolution is $N_r \times N_\phi = 477 \times 1536$, so we also ran this case using a lower resolution (318×1024) and using a higher resolution (636×2048). The final semi-major axis $a_{\rm p,final}$ decreases by 3% (compared to the standard resolution) for the low-resolution run, and increases by 0.5% for the high-resolution run. The numerical results are therefore converged with respect to resolution.

7.4 Discussion

7.4.1 Dependence on Viscosity

The planet is trapped closer to the inner edge of the disk as the viscosity parameter α is increased (see Figure 7.4). This is in qualitative agreement with the results of T11, in which viscous damping reduces the maximum possible enhancement of ILR torques responsible for stopping the migration of the planet. However, the final semi-major axis is relatively insensitive to α for $0.01 \lesssim \alpha \lesssim 0.04$.

T11 estimates the critical viscosity parameter, $\alpha_{\rm crit}$, above which Type I migration can no longer be halted by trapped modes, by considering the effect of viscosity on the resonant gain associated with an individual m-mode, and finds that $\alpha_{\rm crit} \sim h^{5/2} \sim 6 \times 10^{-4}$. Numerically, we find that trapping is possible even for the largest value of α used in our experiments (0.04). We attribute this discrepancy to the fact that modes with multiple values of m can be simultaneously resonant when the planet is trapped (see Figure 7.6), and so α can be much larger than the case of a single mode. This is because, if the ILR torque of a single m-mode were to balance out all of the OLR torques, then it would have to be enhanced by a factof of $\sim h^{-1}$ (approximately the total number of modes contri-

butiong to the total torque), and so a small amount of viscous damping would make this impossible. However, if the ILR torques associated with many values of m collectively cancel out the OLRs, then they must each only be slightly enhanced, and so the cancellation does not depend as sensitively on α

The lowest α we have considered is 4×10^{-4}. For disks with lower viscosities, i.e., "nearly laminar" disks, Type I migration is modified due to the opening of a partial gap and density feedback effects, which may slow down or stop migration (Hourigan & Ward 1984; Ward & Hourigan 1989; Ward 1997; Rafikov 2002; Li et al. 2009; Fung & Chiang 2017).

7.4.2 Dependence on Planet Mass

The semi-major axis at which the planet is trapped decreases as the mass of the planet increases. As suggested by T11, the trapping mechanism is modified when the planet is massive enough for nonlinear damping (due to shocks) to play a role. In particular, when $q \gtrsim (2/3)h^3 \sim 8 \times 10^{-5}$, nonlinear effects are dominant, since waves excited by the planet shock immediately (Goodman & Rafikov 2001). This may set an approximate upper limit for the mass ratios of planets which may be trapped. Indeed, in one of our numerical experiments (not shown), a planet with $q = 10^{-4}$ proceeds to migrate nearly all the way to the inner edge of the disk (the details of this simulation become unreliable at late times, as the distance between the planet and inner edge becomes small and is not resolved by a sufficient number of grid cells). In addition, a sufficiently massive planet may open a gap in the disk, which greatly slows its migration

(Type II migration). This occurs when (Crida et al. 2006)

$$1.08hq^{-1/3} + 50\alpha h^2 q^{-1} \lesssim 1, \tag{7.16}$$

i.e., when $q \gtrsim (10^{-4} - 10^{-3})$, depending on the value of α. Therefore, there may only exist a narrow range of planet masses which result in rapid (Type I) migration which proceeds all the way to the inner edge of the disk.

7.4.3 Dependence on Disk Thickness

Throughout most of this paper, we adopted a disk thickness of $h = H/r = 0.05$, and found that planets can be trapped at about $(1.6 - 3.3)r_{\rm in}$. Additional numerical experiments presented in Figure 7.4 show that for a thicker disk with $h = 0.1$, the final semi-major axis of the planet is at least $2.2r_{\rm in}$, and can be as large as $5.4r_{\rm in}$ for low-viscosity disks. This is significantly larger than the maximum possible trapping radius of about $3r_{\rm in}$ predicted by T11. Nonetheless, for thicker disks, trapping is effective farther from the inner edge, in agreement with T11. This can be understood using the fact that the resonance condition which determines the locations of possible planet traps requires that a small integer number of wavelengths fit between an ILR an the inner edge of the disk, and that the wavelength of density waves increases with h. Therefore, this condition can be satisfied at locations farther from $r_{\rm in}$ for thicker disks. Since we find that the planet is trapped before reaching the inner edge for a plausible range of disk thicknesses, $h = 0.05 - 0.1$, this is unlikely to be a restricting factor for halting migration.

7.4.4 Caveats

Numerical Limitations

The simulations presented in this paper are two-dimensional and employ a locally isothermal sound speed profile and parameterized viscosity. We do not include the effects of magnetic fields, turbulence, and self-gravity, all of which may modify planet migration (see Kley & Nelson 2012; Baruteau et al. 2014). Furthermore, we have implemented a simple boundary condition which is meant to mimic the disk-magnetosphere boundary, which in reality is a complicated, three-dimensional structure. Our simulations can therefore not be compared to state-of-the-art simulations of planet migration (including three-dimensional effects, magnetic fields, radiative transfer, etc.) or star-disk interactions (see Romanova & Owocki 2015). Rather, they serve only to highlight the significant effect of a partially reflective inner disk edge on the migration of a planet, which has until now been neglected in numerical studies. It remains to be seen whether or not this effect persists in more realistic simulations which include a more complete treatment of the relevant physics.

Turbulence

The effective viscosity which we have parameterized using the Shakura-Sunyaev α parameter may be a consequence of turbulence, for example, due to the nonlinear evolution of the magnetorotational instability (MRI; see Balbus & Hawley 1998). Turbulence can significantly alter, and in some cases inhibit Type I migration (e.g., Nelson & Papaloizou 2004). Therefore, it could in principle disrupt the trapping mechanism explored in this paper.

Density fluctuations associated with ideal MRI turbulence lead to stochastic torques with an rms amplitude

$$\Gamma_{\rm turb} \sim \alpha^{1/2} M_{\rm p} \frac{\Sigma a_{\rm p}^2}{M_*} a_{\rm p}^2 \Omega_{\rm p}^2 \tag{7.17}$$

(e.g., Johnson et al. 2006; Okuzumi & Ormel 2013). A planet in a migration trap is stably maintained in the trap by a positive torque that it experiences if it moves inward, whose magnitude is $\sim \Gamma_0$ (equation 7.12; see Figures 7.2–7.3). The planet can therefore stay in the trap provided that $\Gamma_{\rm turb}/\Gamma_0 \lesssim 1$, i.e., if

$$\alpha^{1/2} \left(\frac{H}{r}\right)^2 \lesssim q. \tag{7.18}$$

This is also the approximate criterion which determines if the planet can undergo smooth Type I migration rather than diffusive turbulent migration. For the parameter values we have considered, $\alpha \lesssim 10^{-5}$ is required for the planet to stay in a trap. This places the greatest restriction on the disk properties necessary for the trapping mechanism to be realized in real disks. However, if the anomalous viscosity described by the parameter α is not a result of turbulence, but rather a different physical mechanism, e.g., disk winds resulting from non-ideal magnetohydrodynamical effects (see Turner et al. 2014), then this issue may be circumvented.

7.4.5 Observational Implications

Migration is likely play an important role in the assembly of systems with multiple coplanar, transiting planets, including Kepler 11 (Lissaeur et al. 2011) and TRAPPIST-1 (Gillon et al. 2017). The effect of magnetospheric truncation in halting the migration of planets has been invoked to explain the architecture of such

systems (Mulders et al. 2015; Liu et al. 2017; Lee & Chiang 2017). However, planets are typically assumed to migrate all the way to the inner edge of the disk, or very close to it. Our results indicate that migration may instead be halted at several times the radius of the inner edge, which may have implications for the architecture of these planetary systems. In particular, it may be difficult to explain how planets with orbits close to the stellar corotation radius acquired these orbits, without invoking additional physical mechanisms to bring them inwards.

7.5 Conclusion

We have carried out two-dimensional hydrodynamical simulations of a low-mass planet ($q = 3\times10^{-6} - 3\times10^{-5}$, or $M_{\rm p} = M_\oplus - 10M_\oplus$ for a $1M_*$ star) undergoing Type I migration in a disk with a partially reflective inner edge, which we take to represent the disk-magnetosphere boundary. In agreement with the theory of Tsang 2011 (T11), we find that the migration of the planet can be halted at a semi-major axis several times larger than the inner disk radius $r_{\rm in}$. This occurs due to an enhancement of inner Lindblad torques as a result of a resonance of waves trapped between the orbit of the planet and the inner edge of the disk. A range of azimuthal mode numbers (e.g., $m \approx 5 - 12$) can contribute to the trapping, although this range becomes narrower and tends towards smaller values for more viscous disks. The semi-major axis at which the planet is trapped can be as large as about $3r_{\rm in}$, and generally decreases as either the planet mass or viscosity are increased. For large disk viscosities ($0.01 \lesssim \alpha \lesssim 0.04$), the planet becomes trapped at about $(1.6-2.2)r_{\rm in}$, depending on the disk thickness. This mechanism may play a role in shaping the architecture of systems of short-period planets.

APPENDIX A

WAVE EQUATION AND MODE GROWTH RATES

A.1 Wave Equation Up to First Order in Viscosity and Radial Infall Velocity

In this appendix we summarize the procedure of deriving the equations (2.9)–(2.11). The linearly perturbed continuity equation (2.1) and the Navier-Stokes equation (2.2) in the polar coordinates (r, ϕ) read

$$-\mathrm{i}\tilde{\omega}\delta\Sigma + \frac{1}{r}\partial_r\left(r\Sigma\delta u_r\right) + \frac{\mathrm{i}m}{r}\Sigma\delta u_\phi = -\frac{1}{r}\partial_r\left(ru^r\delta\Sigma\right) \equiv \delta\dot{\Sigma} \tag{A.1}$$

$$-\mathrm{i}\tilde{\omega}\delta u_r - 2\Omega\delta u_\phi + \delta h_{,r} = -\partial_r\left(u_r\delta u_r\right) + \delta\left(\frac{1}{\Sigma}\mathcal{F}_r^{\mathrm{visc}}\right) \equiv \delta\dot{u}_r, \tag{A.2}$$

$$-\mathrm{i}\tilde{\omega}\delta u_\phi + \frac{\kappa^2}{2\Omega}\delta u_r + \frac{\mathrm{i}m}{r}\delta h = -\frac{u_r}{r}\partial_r\left(r\delta u_\phi\right) + \delta\left(\frac{1}{\Sigma}\mathcal{F}_\phi^{\mathrm{visc}}\right) \equiv \delta\dot{u}_\phi, \tag{A.3}$$

The equations were arranged in such a way that the coefficients on the LHS do not depend on the viscosity, while the RHS terms, denoted by $\delta\dot{\Sigma}$, $\delta\dot{u}_r$ and $\delta\dot{u}_\phi$, are at least linear in the viscosity η. The first RHS term in each equation describes the effect of the radial inflow in the stationary case, the second terms in the last two equations are contributions due to the perturbation of the viscous force $\boldsymbol{\mathcal{F}}^{\mathrm{visc}} = \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}$.

In what follows we would like to express the RHS terms $\delta\dot{u}_r$, $\delta\dot{u}_\phi$ and $\delta\dot{\Sigma}$ in terms of the velocity perturbation δu^r, δu^ϕ and enthalpy δh. To simplify the analysis we will further work in the leading order of the WKBJ approximation. Therefore, we assume that the radial wavelength $\lambda_r = 2\pi/k_r$ of the perturbations is much smaller than the radius r. More specifically, we assume that

$k_r r = O(r/H)$, where H is the disk semi-thickness. In thin disks, this assumption is satisfied everywhere except for small regions close to the Lindblad and corotation resonances. In the leading order in the ratio r/H, we obtain

$$\delta\dot{\Sigma} \approx -u_r \frac{\Sigma}{c_s^2}\frac{\mathrm{d}}{\mathrm{d}r}\delta h_{,r} \tag{A.4}$$

$$\delta\dot{u}_r \approx \left[-u_r\frac{\mathrm{d}}{\mathrm{d}r} + \frac{4}{3}\nu\frac{\mathrm{d}^2}{\mathrm{d}r^2}\right]\delta u_r, \tag{A.5}$$

$$\delta\dot{u}_\phi \approx \left[-u_r\frac{\mathrm{d}}{\mathrm{d}r} + \nu\frac{\mathrm{d}^2}{\mathrm{d}r}\right]\delta u_\phi - \nu q A\frac{\Omega}{c_s^2}\frac{\mathrm{d}}{\mathrm{d}r}\delta h. \tag{A.6}$$

In all three equations, the first terms are proportional to u_r and describe the effects of the radial inflow. The second terms in equation (A.5) and (A.6) are the contributions of the perturbed visous force due to velocity field connected to the wave. Finally, the third term of equation (A.6) results from the change of the viscous force due to perturbation of the dynamic viscosity coefficient η.

By substituting expressions (A.5) and (A.6) into the equations (A.2) and (A.3) and expanding the velocity perturbation $\delta\boldsymbol{u}$ in powers of ν,

$$\delta u_r = \delta u_r^0 + \nu\delta u_r^1 + \ldots, \quad \delta u_\phi = \delta u_\phi^0 + \nu\delta u_\phi^1 + \ldots, \tag{A.7}$$

we may solve perturbativelly these equations for $\delta\boldsymbol{u}$ in terms of the enthalpy perturbation δh. Up to the first order in ν (and u_r), we obtain

$$\begin{aligned}
\delta u_r &= \frac{\mathrm{i}}{D}\left(\tilde{\omega}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{2m\Omega}{r}\right)\delta h - \frac{u_r}{D^2}\left(\kappa^2+\tilde{\omega}^2\right)\frac{\mathrm{d}^2}{\mathrm{d}r^2}\delta h + \\
&\quad \frac{\nu}{D^2}\left[\left(\kappa^2+\frac{4}{3}\tilde{\omega}^2\right)\frac{\mathrm{d}^3}{\mathrm{d}r^3} - 2qA\Omega^2\frac{D}{c_s^2}\frac{\mathrm{d}}{\mathrm{d}r}\right]\delta h + O(\nu^2), \\
\delta u_\phi &= \frac{1}{D}\left(\frac{\kappa^2}{2\Omega}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{m\tilde{\omega}}{r}\right)\delta h + \mathrm{i}u_r\frac{\kappa^2\tilde{\omega}}{2\Omega D^2}\frac{\mathrm{d}^2}{\mathrm{d}r^2}\delta h - \\
&\quad \mathrm{i}\tilde{\omega}\frac{\nu}{D^2}\left[\frac{7}{3}\frac{\kappa^2}{2\Omega}\frac{\mathrm{d}^3}{\mathrm{d}r^3} + qA\Omega\frac{D}{c_s^2}\frac{\mathrm{d}}{\mathrm{d}r}\right]\delta h + O(\nu^2).
\end{aligned} \tag{A.8}$$

Finally, by substituting these expressions into the perturbed continuity equation (A.1), we recover the desired equation (2.9) with operators $\hat{L}^0$ and $\hat{L}^1_{\mathrm{in}}$ and $\hat{L}^1_{\mathrm{v}}$.

given by equations (2.5) and (2.10) and (2.11). We remind the reader that these equations are valid only in the leading order in both (H/r)- ratio, radial drift velocity u_r and viscosity ν under the assumpion of short radial wavelength (as compared to the radial coordinate r). Hence, in principle, it cannot be applied in the regions of the Lindblad and corotation resonances. A proper description of those effects would require keeping also the singular terms in equations (A.4)–(A.8).

A.2 Growth Rates of Modes

We would like to find the *global* change of the eigenfrequencies of p-modes trapped in the inner disk. In the following, we will show that it is given by some proper radial average of the Kato (1978) local growth rate over the propagation region. Introducing $\psi = S^{-1/2}\delta h$ with $S = D/(r\Sigma)$, equation (2.9) takes the form

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - V(r)\right]\psi + S^{-1/2}\hat{L}^1\left(S^{1/2}\psi\right) = 0, \tag{A.9}$$

where

$$V(r) = \frac{D}{c_\mathrm{s}^2} + \frac{2m\Omega}{r\tilde{\omega}}\frac{d}{dr}\ln\left(\frac{\Omega\Sigma}{D}\right) + S^{1/2}\frac{d^2}{dr^2}S^{-1/2} + \frac{m^2}{r^2}, \tag{A.10}$$

and $\hat{L}^1$ is either $\hat{L}^1_\mathrm{v}$, or $\hat{L}^1_\mathrm{in}$ or their sum. Assuming that the region of interest does not contain the resonances, the potential $V(r)$ can be approximated by just the first term because the others are by factor of order $(H/r)^2$ smaller. If we express ψ and D as $\psi = \psi_0 + \psi_1$ and $D = D_0 + D_1$ (where ψ_1 and D_1 are of order ν, note that $D_1 = -2\tilde{\omega}_0\omega_1$), we obtain in the zero order equation,

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{D_0}{c_\mathrm{s}^2}\right]\psi_0 = 0, \tag{A.11}$$

and the first order equation,

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{D_0}{c_\mathrm{s}^2}\right]\psi_1 - \frac{D_1}{c_\mathrm{s}^2}\psi_0 + S^{-1/2}\hat{L}^1\left(S^{1/2}\psi_0\right) = 0. \tag{A.12}$$

Multiplying by ψ_0^* and integrating between radii r_1 and r_2 (trapping range of the mode), we obtain

$$\begin{aligned} &2\omega_1 \int_{r_1}^{r_2} \frac{\tilde{\omega}_0}{c_\mathrm{s}^2}\left|\psi_0^2\right|\mathrm{d}r + \int_{r_1}^{r_2}\psi_0^*\left(S^{-1/2}\hat{L}^1 S^{1/2}\psi_0\right)\mathrm{d}r \\ &= -\int_{r_1}^{r_2}\psi_0^*\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{D_0}{c_\mathrm{s}^2}\right]\psi_1 dr \equiv \mathcal{I}. \end{aligned} \tag{A.13}$$

The integral on the RHS can be written as (after two integrations by parts)

$$\left[\psi_1\frac{\mathrm{d}\psi_0^*}{\mathrm{d}r} - \psi_0^*\frac{\mathrm{d}\psi_1}{\mathrm{d}r}\right]_{r_1}^{r_2} - \int_{r_1}^{r_2}\psi_1\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{D_0}{c_\mathrm{s}^2}\right]\psi_0^*\mathrm{d}r. \tag{A.14}$$

Assuming that ω_0 is real, the first integral vanishes. The second term is a factor $O(H/r)$ smaller than the other integrals in equation (A.13). Then, neglecting $\mathcal{I}$, we obtain

$$\omega_1 = -\frac{\int\psi_0^*\left(S^{-1/2}\hat{L}^1 S^{1/2}\psi_0\right)\mathrm{d}r}{2\int\left(\tilde{\omega}_0/c_\mathrm{s}^2\right)|\psi_0|^2\mathrm{d}r}. \tag{A.15}$$

An approximate WKBJ solution of equation (A.11), valid up to the second order, is

$$\begin{aligned} \psi_0 &= \frac{1}{\sqrt{k}}\left[\mathcal{A}_-\exp\left(-\mathrm{i}\int^r k dr\right) + \mathcal{A}_+\exp\left(\mathrm{i}\int^r k dr\right)\right], \\ k &= \sqrt{-D}/c_\mathrm{s}, \end{aligned} \tag{A.16}$$

for which $|\psi_0|^2 = (|\mathcal{A}_-|^2 + |\mathcal{A}_-|^2)/k$. Evaluating the two integrals in equation (A.15) for this solution, we find that

$$\omega_1 = \frac{\int\delta\omega(r)w(r)dr}{\int w(r)dr}, \quad w(r) = \frac{1}{c_\mathrm{s}}\frac{\tilde{\omega}_0}{\sqrt{-D}}. \tag{A.17}$$

Hence, the global change of the frequency of the mode is just the averaged local changes $\delta\omega_\mathrm{v}(r)$ or $\delta\omega_\mathrm{in}(r)$ of equations (2.12) and (2.13) with the weight function $w(r)$.

Although expression (A.17) does not apply close to the Lindblad resonances (LRs), we believe that the first-order LRs can be included as well, because the surrounding regions give negligible contributions to both integrals [$w(r)$ is integrable close to the first-order LRs]. Therefore, the integrals can be extended to the propagation regions of the oscillations. The lower integral however diverges for the second-order LR (at maxima of $\Omega + \kappa/m$) and more sophisticated analysis is needed in those cases.

APPENDIX B

EFFECT OF VISCOSITY ON LINEAR ROSSBY WAVE INSTABILITY

Viscosity plays two main roles in our simulations. It drives the evolution of the disk surface density profile, creating RWI-unstable bumps in the DZ, as well as maintaining the bumps as the RWI attempts to smooth them out. It also has a direct effect on the RWI, damping the growth of the linear instability, as well as affecting its nonlinear evolution. Here we investigate the effect of viscosity on the linear growth rate by performing simulations of the RWI on an artificially-imposed density bump.

We choose a power-law surface density profile modified by a Gaussian bump, given by

$$\Sigma(r) = r^{-1/2}\left\{1 + \chi \exp\left[-\frac{(r - r_{\rm B})^2}{2\sigma^2}\right]\right\}, \tag{B.1}$$

where $r_{\rm B}$ is the location of the bump, σ is its width and χ is its dimensionless amplitude. We henceforth set $r_{\rm B} = 5.5$ and $\chi = 1$, so that σ is the only variable parameter of the bump. The rotation profile $\Omega(r)$ is modified to satisfy centrifugal balance [Eq. (3.6)] given this profile. The resulting vortensity profiles are shown in Fig. B.1. In general, viscosity induces a radial drift velocity given by

$$u_r = \frac{2\Omega}{r^2\kappa^2\Sigma}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^3\nu\Sigma\frac{\mathrm{d}\Omega}{\mathrm{d}r}\right). \tag{B.2}$$

We set $r^3\nu\Sigma\mathrm{d}\Omega/\mathrm{d}r$ to have a constant value in the disk to ensure $u_r = 0$. This is accomplished by choosing a particular $\nu(r)$ profile, given by

$$\nu(r) = \nu_{\rm B}\left(\frac{r}{r_{\rm B}}\right)^{-3}\left[\frac{\Sigma(r)}{\Sigma(r_{\rm B})}\right]^{-1}\left[\frac{\Omega'(r)}{\Omega'(r_{\rm B})}\right]^{-1}, \tag{B.3}$$

where $\Omega' = \mathrm{d}\Omega/\mathrm{d}r$, and $\nu_{\rm B} = \alpha\left(c_s^2/\Omega_{\rm K}\right)_{r_{\rm B}}$, so that the viscosity is described by an effective α at $r_{\rm B}$. Obviously, Eq. (B.3) is not the standard α prescription for

viscosity. This choice of viscosity profile ensures that in a 1D (axisymmetric) simulation, the initial density bump does not diffuse, remaining static on viscous timescales. Therefore in our full 2D simulations, we can test the role of viscosity on the growth of the RWI on a stationary bump. As in our main simulations, the disk has $r_{\rm in} = 1$ and $r_{\rm out} = 12$. For boundary conditions, we choose all fluid variables to be fixed at their initial values at both boundaries. Damping zones are included interior to $r = 2$ and exterior to $r = 10$, in which all variables are relaxed to their initial values on orbital timescales.

The growth rate $\gamma = \mathrm{Im}\,(\omega)$ of an RWI mode (with azimuthal number m) can be written as

$$\gamma = \gamma_0 - \gamma_\nu, \tag{B.4}$$

where $\gamma_0 = \epsilon\Omega$ is the inviscid growth rate (where ϵ depends on the bump profile and sound speed) and γ_ν is the viscous damping rate. We write

$$\gamma_\nu = \beta\frac{\nu}{\sigma^2}, \tag{B.5}$$

where β is of order unity and depends (weakly) on the bump geometry. Near $r_{\rm B}$, we have $\nu \approx \alpha c_s^2/\Omega_{\rm K} = \alpha H^2\Omega$, and so

$$\gamma = \left[\epsilon - \alpha\beta\left(\frac{\sigma}{H}\right)^{-2}\right]\Omega. \tag{B.6}$$

Therefore, we expect the growth of the RWI to be suppressed when $\alpha > \alpha_{\rm crit} = \epsilon\beta^{-1}\,(\sigma/H)^2$.

We simulated the linear growth of the RWI for four bump widths (see Fig. B.1), each with $\alpha = 0$, $\alpha = 0.01$, and $\alpha = 0.025$. For each run, we measure the growth rate of the $m = 4$ mode (shown in Fig. B.2 for $\sigma = H/2$). The growth rates as a function of α are shown in Fig. B.3. For each σ, we find a linear fit to $\gamma(\alpha)$, which determines the value of β, and hence $\alpha_{\rm crit}$, which are given in Table B.1.

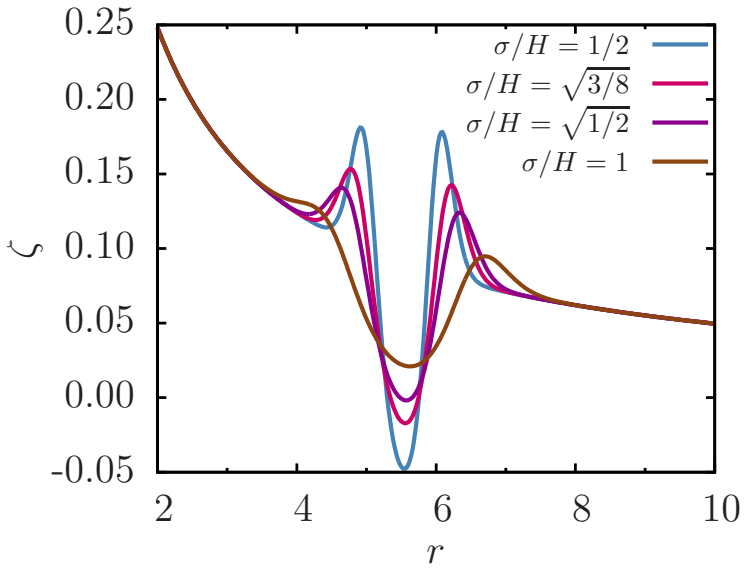

Figure B.1: Vortensity profiles used for simulations of RWI on a fixed density bump. All profiles have $r_B = 5.5$ and $\chi = 1$.

σ/H	β	α_{crit}
$1/2$	1.77	0.0606
$\sqrt{3/8}$	1.74	0.0721
$\sqrt{1/2}$	1.61	0.0751
1	3.71	0.0372

Table B.1: Viscous damping proportionality factor β, and critical viscosity parameter α_{crit}, for different bump widths [see Eq. (B.6)].

For our three narrowest bumps, β is nearly constant (approximately 1.6 – 1.8), and $\alpha_{crit} \approx 0.06 - 0.08$, which is relatively large. For $\sigma = H$, the growth rate is affected more strongly by viscosity (β is twice as large), resulting in $\alpha_{crit} \approx 0.04$. Therefore, the linear RWI may be significantly affected by viscosity for bumps with $\sigma \gtrsim H$ when $\alpha \sim 10^{-2}$.

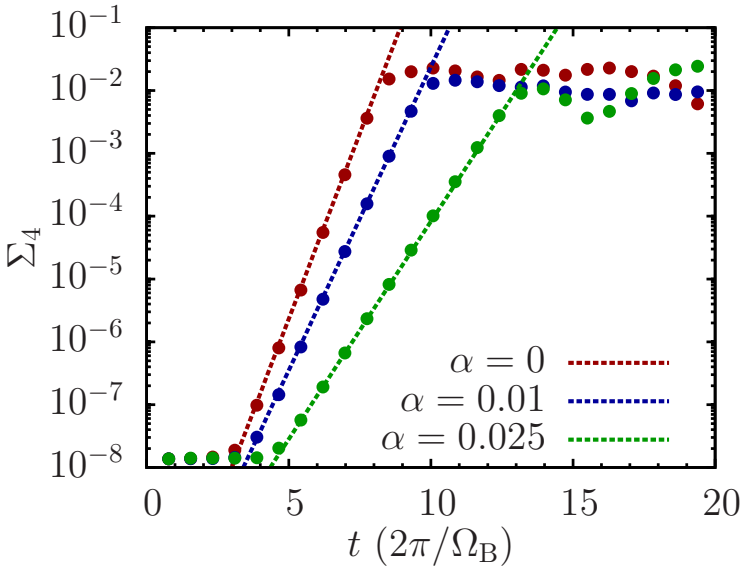

Figure B.2: Evolution of Σ_4 [see Eq. (3.14)] for $\sigma = H/2$ and several values of α, demonstrating viscous damping of RWI growth. The dashed lines are fits to the linear growth phase, from which the growth rates are determined.

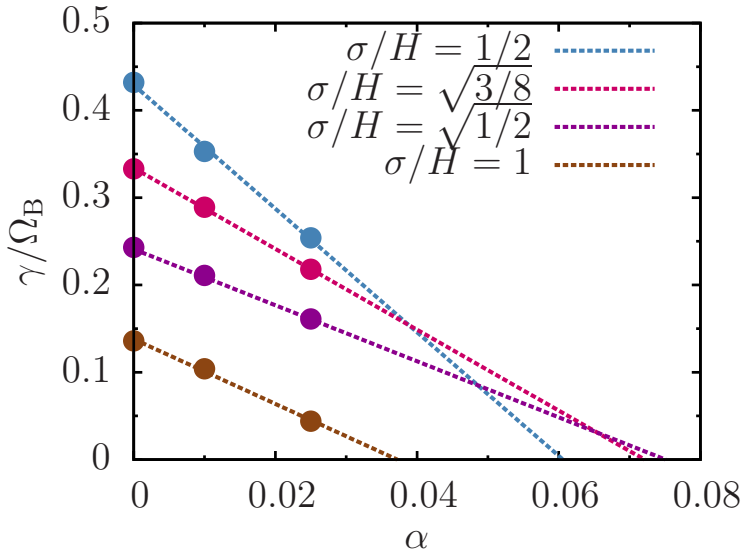

Figure B.3: Growth rates as as a function of α (points) for different bump widths. The dashed lines are linear fits to the points, whose x-intercepts give an estimate of $\alpha_{\rm crit}$.

APPENDIX C

EXPANSION OF BINARY POTENTIAL AS POWER SERIES IN ECCENTRICITY

We expand $(r_{12}/a)^{-(l+1)}\cos\left(m\phi - \mu\phi_2'\right)$ [for a circumstellar disk, see equation (5.10)], or $(r_{12}/a)^{l}\cos\left(m\phi - \mu\phi_2'\right)$ [for a circumbinary disk, see equation (5.15)], as a power series in e. Here r_{12} is the binary separation and ϕ_2' specifies the azimuthal position of M_2 relative to M_1 in the orbital plane of the binary, which, without loss of generality, can be chosen to be the true anomaly of the binary. We use the elliptic expansions (e.g., Murray & Dermott 1999; Brouwer & Clemence 1961) of r_{12}/a and ϕ_2', in terms of e and mean anomaly $M = \Omega_{\rm B} t$. For the binary separation r_{12} we have

$$\begin{aligned}\frac{r_{12}}{a} &= 1 + \frac{1}{2}e^2 - 2e\sum_{s=1}^{\infty}\frac{1}{s^2}\frac{\mathrm{d}}{\mathrm{d}e}J_s(se)\cos(sM)\\ &= 1 - e\cos(M) + \frac{e^2}{2}\left[1-\cos(2M)\right] + \frac{3e^3}{8}\left[\cos(M)-\cos(3M)\right]\\ &+ \frac{e^4}{3}\left[\cos(2M)-\cos(4M)\right] + O\left(e^5\right).\end{aligned} \tag{C.1}$$

The following expression relating the mean anomaly M and eccentric anomaly E is useful:

$$\begin{aligned}E &= M + 2\sum_{s=1}^{\infty}\frac{1}{s}J_s(se)\sin(sM)\\ &= M + e\sin(M) + \frac{e^2}{2}\sin(2M) + e^3\left[\frac{3}{8}\sin(3M) - \frac{1}{8}\sin(M)\right]\\ &+ e^4\left[\frac{1}{3}\sin(4M) - \frac{1}{6}\sin(2M)\right] + O\left(e^5\right).\end{aligned} \tag{C.2}$$

The true anomaly is given by

$$\begin{aligned}\phi_2' &= \int_0^M \frac{\left(1-e^2\right)^{\frac{1}{2}}}{\left[1-e\cos\left(E\right)\right]^2}\mathrm{d}M \\ &= M + 2e\sin\left(M\right) + \frac{5e^2}{4}\sin\left(2M\right) + e^3\left[\frac{13}{12}\sin\left(3M\right) - \frac{1}{4}\sin\left(M\right)\right] \\ &+ e^4\left[\frac{103}{96}\sin\left(4M\right) - \frac{11}{24}\sin\left(2M\right)\right] + O\left(e^5\right).\end{aligned} \tag{C.3}$$

Using the above expressions we can obtain the power series (in e) for $(r_{12}/a)^{-(l+1)}\cos\left(m\phi - \mu\phi_2'\right)$ and $(r_{12}/a)^{l}\cos\left(m\phi - \mu\phi_2'\right)$, and gathering $\cos\left[m\phi - (\mu+n)\,\Omega_{\mathrm{B}}t\right]$ terms [equations (5.10) and (5.15)], we find the coefficients $C^{\mathrm{CS}}_{l,\mu,n}$ and $C^{\mathrm{CB}}_{l,\mu,n}$.

Using computer algebra, we have computed $C^{\mathrm{CS}}_{l,\mu,n}$ and $C^{\mathrm{CB}}_{l,\mu,n}$ for $|n| \leq 8$. Each coefficient is proportional, to leading order, to $e^{|n|}$, with higher order terms proportional to $e^{|n|+2}$, $e^{|n|+4}$, and so on. We have included terms up to order e^{10}, so that every $C_{l,\mu,n}$ includes at least one higher order correction in eccentricity after its leading order. The leading terms for $|n| \leq 4$ are shown below.

$$\begin{aligned}C^{\mathrm{CS}}_{l,\mu,0} &= 1 \\ C^{\mathrm{CS}}_{l,\mu,\pm1} &= \frac{1}{2}e\left[l \pm 2\mu + 1\right] \\ C^{\mathrm{CS}}_{l,\mu,\pm2} &= \frac{1}{8}e^2\left[l^2 + (5 \pm 4\mu)l + 4\mu^2 \pm 9\mu + 4\right] \\ C^{\mathrm{CS}}_{l,\mu,\pm3} &= \frac{1}{48}e^3\left[l^3 + 6(2 \pm \mu)l^2 + \left(12\mu^2 \pm 45\mu + 38\right)l\right. \\ &\quad \left.\pm 8\mu^3 + 42\mu^2 \pm 65\mu + 27\right] \\ C^{\mathrm{CS}}_{l,\mu,\pm4} &= \frac{1}{384}e^4\left[l^4 + (22 \pm 8\mu)l^3 + \left(24\mu^2 \pm 126\mu + 155\right)l^2\right. \\ &\quad + \left(\pm 32\mu^3 + 240\mu^2 \pm 558\mu + 390\right)l + 16\mu^4 \pm 152\mu^3 \\ &\quad \left.+ 499\mu^2 \pm 646\mu + 256\right]\end{aligned} \tag{C.4}$$

$$
\begin{aligned}
C^{\mathrm{CB}}_{l,\mu,0} &= 1 \\
C^{\mathrm{CB}}_{l,\mu,\pm1} &= \frac{1}{2} e \left[-l \pm 2\mu\right] \\
C^{\mathrm{CB}}_{l,\mu,\pm2} &= \frac{1}{8} e^2 \left[l^2 - (3 \pm 4\mu) l + \mu(4\mu \pm 5)\right] \\
C^{\mathrm{CB}}_{l,\mu,\pm3} &= \frac{1}{48} e^3 \left[-l^3 + 3(3 \pm 2\mu) l^2 - \left(12\mu^2 \pm 33\mu + 17\right) l \right. \\
&\quad \left. + 2\mu \left(\pm 4\mu^2 + 15\mu \pm 13\right)\right] \\
C^{\mathrm{CB}}_{l,\mu,\pm4} &= \frac{1}{384} e^4 \left[l^4 - 2(9 \pm 4\mu) l^3 + \left(24\mu^2 \pm 102\mu + 95\right) l^2 \right. \\
&\quad - 2 \left(\pm 16\mu^3 + 96\mu^2 \pm 165\mu + 71\right) l \\
&\quad \left. + \mu \left(16\mu^3 \pm 120\mu^2 + 283\mu \pm 206\right)\right]
\end{aligned}
\tag{C.5}
$$

Note that these coefficients are equivalent to the Hansen coefficients $X_c^{a,b}(e)$ (Murray & Dermott 1999), with

$$
C^{\mathrm{CS}}_{l,\mu,n} = X^{-(l+1),\mu}_{\mu+n}(e) \quad \text{and} \quad C^{\mathrm{CB}}_{l,\mu,n} = X^{l,\mu}_{\mu+n}(e). \tag{C.6}
$$

The Hansen coefficients are defined according to

$$
\left[\frac{r_{12}(t)}{a}\right]^{\lambda} \exp\left[-\mathrm{i}\mu\phi_2'(t)\right] = \sum_N X^{\lambda,\mu}_N \exp\left[-\mathrm{i}N\Omega_{\mathrm{B}} t\right], \tag{C.7}
$$

which can be inverted, giving the following expression:

$$
X^{\lambda,\mu}_N(e) = \frac{1}{\pi} \int_0^{\pi} \frac{\cos\left\{N\left[E - e\sin(E)\right] - \mu\phi_2'\right\}}{\left[1 - e\cos(E)\right]^{-(\lambda+1)}} \mathrm{d}E. \tag{C.8}
$$

They can also be computed recursively using the Newcomb operators.

APPENDIX D

ANGULAR MOMENTUM CONSERVATION IN CIRCUMBINARY DISKS

The equation of motion for the gas in the disk is

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\,\mathbf{u} = -\frac{1}{\Sigma}\nabla P - \nabla\Phi + \mathbf{f}_{\rm visc}, \tag{D.1}$$

where Φ is the gravitational potential of the binary and $\mathbf{f}_{\rm visc}$ is the viscous force per unit mass. Starting with the ϕ component,

$$\frac{\partial u_\phi}{\partial t} + u_r\frac{\partial u_\phi}{\partial r} + \frac{u_\phi}{r}\left(u_r + \frac{\partial u_\phi}{\partial \phi}\right) = -\frac{1}{r\Sigma}\frac{\partial P}{\partial \phi} - \frac{1}{r}\frac{\partial \Phi}{\partial r} + f_{{\rm visc},\phi}, \tag{D.2}$$

multiplying by r, and defining the angular momentum per unit mass $l = ru_\phi$, we have

$$\frac{\partial l}{\partial t} + u_r\frac{\partial l}{\partial r} + \frac{u_\phi}{r}\frac{\partial l}{\partial \phi} = -\frac{1}{\Sigma}\frac{\partial P}{\partial \phi} - \frac{\partial \Phi}{\partial \phi} + t_z^{\rm visc}, \tag{D.3}$$

where $t_z^{\rm visc} = r f_{{\rm visc},\phi}$ is the viscous torque per unit mass. Noting that left-hand side of this equation is equal to $\partial l/\partial t + (\mathbf{u}\cdot\nabla)l = {\rm d}l/{\rm d}t$, we see that this is the evolution equation for l. Next, multiplying Eq. (D.3) by $r\Sigma$, making use of the identity

$$\Sigma\frac{\partial l}{\partial t} = \frac{\partial}{\partial t}(\Sigma l) - l\frac{\partial \Sigma}{\partial t}, \tag{D.4}$$

and the continuity equation,

$$\frac{\partial \Sigma}{\partial t} = -\nabla\cdot(\Sigma\mathbf{u}), \tag{D.5}$$

we arrive at

$$\frac{\partial}{\partial t}(r\Sigma l) + \frac{\partial}{\partial r}(r\Sigma u_r l) + \frac{\partial}{\partial \phi}\left(\Sigma u_\phi l\right) = -r\frac{\partial P}{\partial \phi} - r\Sigma\frac{\partial \Phi}{\partial \phi} + r\Sigma t_z^{\rm visc}. \tag{D.6}$$

Integrating in ϕ, we then have

$$\frac{\partial}{\partial t}\left(\frac{{\rm d}J}{{\rm d}r}\right) = \frac{\partial \dot{J}_{\rm adv}}{\partial r} + \frac{{\rm d}T_{\rm grav}}{{\rm d}r} - \frac{\partial \dot{J}_{\rm visc}}{\partial r}, \tag{D.7}$$

where we have defined

$$\frac{\mathrm{d}J(r,t)}{\mathrm{d}r} = \oint r\Sigma l \mathrm{d}\phi, \tag{D.8}$$

$$\dot{J}_{\mathrm{adv}}(r,t) = -\oint r\Sigma u_r l \mathrm{d}\phi, \tag{D.9}$$

$$\frac{\mathrm{d}T_{\mathrm{grav}}(r,t)}{\mathrm{d}r} = -\oint r\Sigma \frac{\partial\Phi}{\partial\phi}\mathrm{d}\phi, \tag{D.10}$$

and

$$\frac{\partial\dot{J}_{\mathrm{visc}}(r,t)}{\partial r} = -\oint r\Sigma t_z^{\mathrm{visc}}\mathrm{d}\phi, \tag{D.11}$$

which are the angular momentum per unit r, the inward flux of angular momentum due to advection, the gravitational torque per unit r, and the viscous torque per unit r. Eq. (D.7) describes the rate of change of angular momentum in a ring of unit width. In general, the viscous torque is given by

$$\dot{J}_{\mathrm{visc}} = -\oint r^3 \nu\Sigma\left[\frac{\partial}{\partial r}\left(\frac{u_\phi}{r}\right) + \frac{1}{r^2}\frac{\partial u_{\mathrm{r}}}{\partial\phi}\right]\mathrm{d}\phi. \tag{D.12}$$

Taking the time average of Eq. (D.7), and assuming a steady state ($\partial/\partial t \to 0$), we have

$$\frac{\partial}{\partial r}\left(\langle\dot{J}_{\mathrm{adv}}\rangle - \langle\dot{J}_{\mathrm{visc}}\rangle\right) + \left\langle\frac{\mathrm{d}T_{\mathrm{grav}}}{\mathrm{d}r}\right\rangle = 0. \tag{D.13}$$

Finally, integrating radially from r to r_{out}, and noting that $\mathrm{d}T_{\mathrm{grav}}/\mathrm{d}r$ vanishes far from the binary, we find that

$$\langle\dot{J}\rangle_r = \langle\dot{J}_{\mathrm{adv}}\rangle_{r_{\mathrm{out}}} - \langle\dot{J}_{\mathrm{visc}}\rangle_{r_{\mathrm{out}}}. \tag{D.14}$$

We have defined

$$\langle\dot{J}\rangle_r \equiv \langle\dot{J}_{\mathrm{adv}}\rangle_r - \langle\dot{J}_{\mathrm{visc}}\rangle_r - \langle T_{\mathrm{grav}}^{>r}\rangle, \tag{D.15}$$

where

$$\langle T_{\mathrm{grav}}^{>r}\rangle \equiv \int_r^{r_{\mathrm{out}}}\left\langle\frac{\mathrm{d}T_{\mathrm{grav}}}{\mathrm{d}r}\right\rangle\mathrm{d}r. \tag{D.16}$$

Eq. (D.15) gives the net angular momentum flux through the disk at radius r, including contributions from mass advection, viscous stress and gravitational

torque. We also define its non-averaged analogue $\dot{J}$, by dropping the time averages. Eq. (D.14) indicates that $\langle \dot{J} \rangle$ is a constant across the disk.

Since $\dot{J}_{\rm visc} = 0$ at the (physical) inner edge of the disk, we find that $\langle \dot{J} \rangle_{r_{\rm in}} = \langle \dot{J}_{\rm adv} \rangle_{r_{\rm in}} - \langle T^{\rm tot}_{\rm grav} \rangle$ (where $\langle T^{\rm tot}_{\rm grav} \rangle$ is the gravitational torque exerted on the entire disk) is the net angular momentum transfer rate to the binary, which we denote $\langle \dot{J}_{\rm B} \rangle$. Therefore,

$$\langle \dot{J} \rangle_r = \langle \dot{J}_{\rm B} \rangle \tag{D.17}$$

for all r. In a steady state, the mass accretion rate $\dot{M}$ is constant throughout the disk, therefore, we can write

$$\dot{J}_{\rm B} = \dot{M} l_0. \tag{D.18}$$

The specific angular momentum l_0 represents the net angular momentum received by the binary per unit mass it accretes from the disk.

BIBLIOGRAPHY

[1] Abramowicz, M. A., Kluźniak, W., 2001, A&A, 374, L19

[2] Abramowicz, M. A., Kluźniak, W., Bursa, M., Horák, J., Rebusco, P., Török, G., 2007, RevMexAA, 27, 8

[3] Armitage, P. J., Natarajan, P., 2002, ApJL, 567, L9

[4] Armitage, P. J., 2011, ARA&A, 49, 195

[5] Armitage, P. J., Livio, M., Pringle, J. E., 2001, MNRAS, 324, 705

[6] Artymowicz, P., Lubow, S. H., 1994, ApJ, 421, 651

[7] Artymowicz, P., Lubow, S. H., 1996, ApJL, 467, L77

[8] Bae, J., Hartmann, L., Zhu, Z., Nelson, R. P., 2014, ApJ, 795, 61

[9] Bai, X.-N., 2013, ApJ, 772, 96

[10] Bai, X.-N. 2014, ApJ, 791, 137

[11] Bai, X.-N., 2014b, ApJ, 798, 84

[12] Bai, X.-N. 2016, ApJ, 821, 80

[13] Bai, X.-N., & Stone, J. M. 2013, ApJ, 769, 76

[14] Balbus, S. A., Hawley, J. F., 1991, ApJ, 376, 214

[15] Balbus, S. A., Hawley, J. F., 1998, Rev. Mod. Phys., 70, 1

[16] Balbus, S. A., Papaloizou, J. C. B., 1999, ApJ, 521, 650

[17] Barge, P., Sommeria, J., 1995, A&A, 295, L1

[18] Baruteau, C., & Masset, F. 2008, ApJ, 678, 483

[19] Baruteau, C., Crida, A., Paardekooper, S.-J., et al. 2014, Protostars and Planets VI, 667

[20] Baruteau, C., & Zhu, Z. 2016, MNRAS, 458, 3927

[21] Bary, J. S., Petersen, M. S., 2014, ApJ, 792, 64

[22] Bate M. R., Bonnell I. A., Bromm V., 2003, MNRAS, 339, 577

[23] Begelman, M. C., Blandford, R. D., Rees, M. J., 1980, Nature, 287, 307

[24] Belloni, T. M., Sanna, A., Méndez, M., 2012, MNRAS, 426, 1701

[25] Benítez-Llambay, P., & Masset, F. S. 2016, ApJS, 223, 11

[26] Birnstiel, T., Fang, M., & Johansen, A. 2016, Space Sci. Rev.,

[27] Bisnovatyi-Kogan, G. S., Ruzmaikin, A. A., 1974, Ap&SS, 28, 45

[28] Bisnovatyi-Kogan, G. S., Ruzmaikin, A. A., 1976, Ap&SS, 42, 401

[29] Bitsch, B., Morbidelli, A., Lega, E., Kretke, K., Crida, A., 2014, A&A, 570, A75

[30] Blaes, O. M., Arras, P., Fragile, P. C., 2006, MNRAS, 369, 1235

[31] Brouwer, D., Clemence, G. M., 1961, *Methods of Celestial Mechanics*, Academic Press

[32] Bruderer, S., van der Marel, N., van Dishoeck, E. F., & van Kempen, T. A. 2014, A&A, 562, A26

[33] Capelo, H. L., Herbst, W., Leggett, S. K., Hamilton, C. M., Johnson, J. A., 2012, ApJ, 757, L18

[34] Casassus, S., van der Plas, G., M, S. P., et al. 2013, Nature, 493, 191

[35] Chan, C.-K., 2009, ApJ, 704, 68

[36] Chang, P., Oishi, J. S., 2010, ApJ, 721, 1593

[37] Chang, P., Strubbe, L. E., Menou, K., Quataert, E., 2010, MNRAS, 407, 2007

[38] Chapon, D., Mayer, L., Teyssier, R., 2013, MNRAS, 429, 3114

[39] Chauvin, G., Beust, H., Lagrange, A.-M., Eggenberger, A., 2011, A&A, 528, A8

[40] Chiang, E. I., Murray-Clay, R. A., 2004, ApJ, 607, 913

[41] Chiang, E., Youdin, A. N., 2010, AREPS, 38, 493

[42] Chiang, E., & Laughlin, G. 2013, MNRAS, 431, 3444

[43] Cleeves, L. I., Adams, F. C., & Bergin, E. A. 2013, ApJ, 772, 5

[44] Colella, P., Woodward, P. R., 1984, Journal of Computational Physics, 54, 174

[45] Crida, A., Morbidelli, A., & Masset, F. 2006, Icarus, 181, 587

[46] Cuadra, J., Armitage, P. J., Alexander, R. D., Begelman, M. C., 2009, MNRAS, 393, 1423

[47] D'Angelo, G., & Lubow, S. H. 2010, ApJ, 724, 730

[48] Davis, C. J., Mundt, R., Eislóffel, J., 1994, ApJ, 437, L55

[49] Desch, S. J., Turner, N. J., 2015, ApJ, 811, 156

[50] de Val-Borro, M. et al., 2006, MNRAS, 370, 529

[51] de Val-Borro, M., Gahm, G. F., Stempels, H. C., Pepliński, A., 2011, MNRAS, 413, 2679

[52] Dolence, J. C., Gammie, C. F., Shiokawa, H., Noble, S. C., 2012, ApJ, 746, L10

[53] Done, C., Gierliński, M., Kubota, A., 2007, A&AR, 15, 1

[54] D'Orazio, D. J., Haiman, Z., MacFadyen, A., 2013, MNRAS, 436, 2997

[55] D'Orazio, D. J., Haiman, Z., Duffell, P., Farris, B. D., MacFadyen, A. I., 2015, MNRAS, 452, 2540

[56] D'Orazio, D. J., Haiman, Z., Duffell, P., MacFadyen, A., Farris, B., 2016, MNRAS, 459, 2379

[57] Dotti, M., Colpi, M., Haardt, F., Mayer, L., 2007, MNRAS, 379, 956

[58] Doyle, L. R., et al., 2011, Science, 333, 1602

[59] Duffell, P. C., MacFadyen, A. I., 2013, ApJ, 769, 41

[60] Duffell, P. C., 2016, ApJS, 226, 2

[61] Dullemond, C. P. 2012, Astrophysics Source Code Library, ascl:1202.015

[62] Dumusque, X., et al., 2012 Nature, 491, 207

[63] Dunhill, A. C., Cuadra, J., Dougados, C., 2015, MNRAS, 448, 3545

[64] Dutrey, A., Guilloteau, S., Simon, M., 1994, A&A, 286, 149

[65] Eggleton, P. P., 1983, ApJ, 268, 368

[66] Escala, A., Larson, R. B., Coppi, P. S., Mardones, D., 2005, ApJ, 630, 152

[67] Espaillat, C., Muzerolle, J., Najita, J., et al. 2014, Protostars and Planets VI, 497

[68] Farris, B. D., Duffell, P., MacFadyen, A. I., Haiman, Z., 2014, ApJ, 783, 134

[69] Fielding, D. B., McKee, C. F., Socrates, A., Cunningham, A. J., Klein, R. I., 2015, MNRAS, 450, 3306

[70] Fleming, D. P., Quinn, T. R., 2017, MNRAS, 464, 3343

[71] Fleming, T., Stone, J. M., 2003, ApJ, 585, 908

[72] Follette, K. B., Grady, C. A., Swearingen, J. R., et al. 2015, ApJ, 798, 132

[73] Ford, E. B., Kozinsky, B., Rasio, F. A, 2000, ApJ, 535, 385

[74] Foucart, F., Lai, D., 2014, MNRAS, 445, 1731

[75] Fu, W., Lai, D., 2009, ApJ, 690, 1386

[76] Fu, W., Lai, D., 2011, MNRAS, 410, 399

[77] Fu, W., Lai, D., 2012, MNRAS, 423, 831

[78] Fu, W., Lai, D., 2013, MNRAS, 431, 3697

[79] Fu, W., Li, H., Lubow, S., & Li, S. 2014, ApJL, 788, L41

[80] Fu, W., Li, H., Lubow, S., Li, S., & Liang, E. 2014, ApJL, 795, L39

[81] Fung, J., & Chiang, E. 2017, ApJ, 839, 100

[82] Gammie, C. F. 1996, ApJ, 457, 355

[83] Gholipour, M., Nejad-Asghar, M., 2014, MNRAS, 441, 1910

[84] Ghosh, P., & Lamb, F. K. 1979, ApJ, 232, 259

[85] Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Nature, 542, 456

[86] Godon, P., Livio, M., 1999, ApJ, 523, 350

[87] Godon, P., Livio, M., 2000, ApJ, 537, 396

[88] Goldreich, P., Tremaine, S., 1979, ApJ, 233, 857

[89] Goldreich, P., Tremaine, S., 1980, ApJ, 241, 425

[90] Goldreich, P., Nicholson, P. D., 1989, ApJ, 342, 1075

[91] Goodchild, S., Ogilvie, G. 2006, MNRAS, 368, 1123

[92] Goodman, J., Rafikov, R. R., 2001, ApJ, 552, 793

[93] Guñther, R., Kley, W. 2002, A&A, 387, 550

[94] Haiman, Z., Kocsis, B., Menou, K., 2009, ApJ, 700, 1952

[95] Haisch, K. E., Jr., Lada, E. A., Lada, C. J., 2001, ApJL, 553, L153

[96] Hanawa, T., Ochi, Y., Ando, K., 2010, ApJ, 708, 485

[97] Harrington, R. S., 1968, AJ, 73, 190

[98] Hartmann, L., Calvet, N., Gullbring, E., D'Alessio, P, 1998, ApJ, 495, 385

[99] Hatzes, A. P., Cochran, W. D., Endl, M., McArthur, B., Paulson, D. B., Walker, G. A. H., Campbell, B., Yang, S., 2003, ApJ, 599, 1383

[100] Hayashi, C. 1981, Progress of Theoretical Physics Supplement, 70, 35

[101] Henisey, K. B., Blaes, O. M., Fragile, P. C., Ferreira, B. T., 2009, ApJ, 706, 705

[102] Hioki, T., Itoh, Y., Oasa, Y., Fukagawa, M., Hayashi, M., 2011, PASJ, 63, 543

[103] Holman, M. J., Wiegert, P. A., 1999, AJ, 117, 621

[104] Horák, J., Lai, D., 2013, MNRAS, 434, 2761

[105] Hourigan, K., & Ward, W. R. 1984, Icarus, 60, 29

[106] Igumenshchev, I. V., Narayan, R., Abramowicz, M. A., 2003, ApJ, 592, 1042

[107] Isella, A., Carpenter, J. M., & Sargent, A. I. 2009, ApJ, 701, 260

[108] Isella, A., Pérez, L. M., Carpenter, J. M., et al. 2013, ApJ, 775, 30

[109] Ivanov, P. B., Papaloizou, J. C. B., Polnarev, A. G., 1999, MNRAS, 307, 79

[110] Jaffe, A. H., Backer, D. C. 2003, ApJ, 583, 616

[111] Jang-Condell, H., 2015, ApJ, 799, 147

[112] Jensen, E. L. N., Dhital, S., Stassun, K. G., Patience, J., Herbst, W., Walter, F. M., Simon, M., Basri, G., 2007, AJ, 134, 241

[113] Jensen, E. L. N., Akeson, R., 2014, Nature, 511, 567

[114] Jin, S., Li, S., Isella, A., Li, H., & Ji, J. 2016, ApJ, 818, 76

[115] Johnson, E. T., Goodman, J., & Menou, K. 2006, ApJ, 647, 1413

[116] Kato, S., 1978, MNRAS, 185, 629

[117] Kato, S., Honma, F. & Matsumoto, R. 1988a, MNRAS, 231, 37

[118] Kato, S., Honma, F. & Matsumoto, R. 1988b, PASJ, 40, 709

[119] Kato, S., Wu, X.-B., Yang, L.-T. & Yang, Z.-L. 1993, MNRAS, 260, 317

[120] Kato, S., 2001, PASJ, 53, 1

[121] Kelley, L. Z., Blecha, L., & Hernquist, L. 2017, MNRAS, 464, 3131

[122] Kennedy, G. M., Wyatt, M. C., Sibthorpe, B., Duchêne, G., Kalas, P., Matthews, B. C., Greaves, J. S., Su, K. Y. L., Fitzgerald, M. P., 2012a, MNRAS, 421, 2264

[123] Kennedy, G. M., Wyatt, M. C., Sibthorpe, B., Phillips, N. M., Matthews, B. C., Greaves, J. S., 2012b, MNRAS, 426, 2115

[124] Klahr, H. H., Bodenheimer, P., 2003, ApJ, 582, 869

[125] Klessen R. S., 2011, EAS Pub. Series, 51, 133

[126] Kley, W., Haghighipour, N., 2014, A&A, 564, A72

[127] Kley, W., Haghighipour, N., 2015, A&A, 581, A20

[128] Kley, W., & Nelson, R. P. 2012, ARA&A, 50, 211

[129] Koenigl, A. 1991, ApJL, 370, L39

[130] Kretke, K. A., Lin, D. N. C., 2007, ApJL, 664, L55

[131] Lai, D., Tsang, D., 2009, MNRAS, 393, 979

[132] Lai, D., Fu, W., Tsang, D., Horák, J., Yu, C., 2013, IAUS, 290, 57

[133] Latter, H. N., Ogilvie, G. I., 2010, Icarus, 210, 318

[134] Lee, E. J., & Chiang, E. 2017, ApJ, 842, 40

[135] Lesur, G., Kunz, M. W., & Fromang, S. 2014, A&A, 566, A56

[136] Li, H., Finn, J. M., Lovelace, R. V. E., & Colgate, S. A. 2000, ApJ, 533, 1023

[137] Li, H., Colgate, S. A., Wendroff, B., & Liska, R. 2001, ApJ, 551, 874

[138] Li, H., Li, S., Koller, J., et al. 2005, ApJ, 624, 1003

[139] Li, H., Lubow, S. H., Li, S., & Lin, D. N. C. 2009, ApJL, 690, L52

[140] Lin, D. N. C., Papaloizou, J., 1979, MNRAS, 186, 799

[141] Lin, D. N. C., Papaloizou, J., 1986, ApJ, 309, 846

[142] Lin, M.-K., 2014, MNRAS, 437, 575

[143] Lines, S., Leinhardt, Z. M., Baruteau, C., Paardekooper, S.-J., Carter, P. J., 2015, A&A, 582, A5

[144] Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nature, 470, 53

[145] Liu, B., Muñoz, D. J., Lai, D., 2015a, MNRAS, 447, 747

[146] Liu, B., Lai, D., Yuan, Y.-F., 2015b, Phys. Rev. D, 92, 124048

[147] Liu, B., Ormel, C. W., & Lin, D. N. C. 2017, A&A, 601, A15

[148] Lodato, G., Rice, W. K. M., 2004, MNRAS, 351, 630

[149] Long, M., Romanova, M. M., & Lovelace, R. V. E. 2005, ApJ, 634, 1214

[150] Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, ApJ, 513, 805

[151] Lovelace, R. V. E., Hohlfeld, R. G., 2013, MNRAS, 429, 529

[152] Lovelace, R. V. E., Romanova, M. M., 2014, Fluid Dyn. Res., 46, 041401

[153] Lubow, S. H., 1991, ApJ, 381, 259

[154] Lubow, S. H., Artymowicz, P., 2000, Protostars and Planets IV, 731

[155] Lubow, S. H., Martin, R. G., Nixon, C., 2015, ApJ, 800, 96

[156] Lynden-Bell, D., & Pringle, J. E. 1974, MNRAS, 168, 603

[157] Lyra, W., Johansen, A., Zsom, A., Klahr, H., & Piskunov, N. 2009, A&A, 497, 869

[158] Lyra, W., & Mac Low, M.-M. 2012, ApJ, 756, 62

[159] Lyra, W., & Lin, M.-K. 2013, ApJ, 775, 17

[160] Lyra, W., Turner, N. J., & McNally, C. P. 2015, A&A, 574, A10

[161] Lyubarskij, Y. E., Postnov, K. A., Prokhorov, M. E., 1994, MNRAS, 266, 583

[162] MacFadyen, A. I., Milosavljevíc, M., 2008, ApJ, 672, 83

[163] Mao, S. A., Psaltis, D., Milsom, J. A., 2009, ApJ, 703, 717

[164] Mardling, R. A., Lin, D. N. C., 2002, ApJ, 573, 829

[165] Martin, R. G., Lubow, S. H., 2011, ApJ, 740, L6

[166] Martin, R. G., Lubow, S. H., 2014, MNRAS, 437, 682

[167] Masset, F., 2000, A&AS, 141, 165

[168] Masset, F. S., & Papaloizou, J. C. B. 2003, ApJ, 588, 494

[169] Mathieu, R. D., Stassun, K., Basri, G., et al., 1997, AJ, 113, 1841

[170] Mayer, L., Kazantzidis, S., Madau, P., et al., 2007, Science, 316, 1874

[171] McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

[172] McKinney, J. C., Tchekhovskoy, A., Blandford, R. D., 2012, MNRAS, 423, 3083

[173] Méheut, H., Casse, F., Varnière, P., Tagger, M., 2010, A&A 516, 31

[174] Méheut, H., Keppens, R., Casse, F., & Benz, W. 2012, A&A, 542, A9

[175] Méheut, H., Meliani, Z., Varniere, P., & Benz, W. 2012, A&A, 545, A134

[176] Méheut, H, Yu, C., Lai, D., 2012, MNRAS, 422, 2399

[177] Méheut, H., Lovelace, R. V. E., Lai, D., 2013, MNRAS, 430, 1988

[178] Meschiari, S., 2012, ApJ, 752, 71

[179] Meyer-Vernet, N., Sicardy, B., 1987, Icarus, 69, 157

[180] Mignone, A., Bodo, G., Massaglia, S., Matsakos, T., Tesileanu, O., Zanni, C., Ferrari, A., 2007, ApJS, 170, 228

[181] Milosavljević, M., Phinney, E. S., 2005, ApJL, 622, L93

[182] Milsom, J. A., Taam, R. E., 1996, MNRAS, 283, 919

[183] Miranda, R., Horák, J., & Lai, D. 2015, MNRAS, 446, 240

[184] Miranda, R., & Lai, D. 2015, MNRAS, 452, 2396

[185] Miranda, R., Lai, D., & Méheut, H. 2016, MNRAS, 457, 1944

[186] Miranda, R., Li, H., Li, S., & Jin, S. 2017, ApJ, 835, 118

[187] Miranda, R., Muñoz, D. J., & Lai, D. 2017, MNRAS, 466, 1170

[188] Mittal, T., & Chiang, E. 2015, ApJL, 798, L25

[189] Moriwaki, K., Nakagawa, Y., 2004, ApJ, 609, 1065

[190] Mulders, G. D., Pascucci, I., & Apai, D. 2015, ApJ, 798, 112

[191] Muñoz, D. J., Kratter, K., Springel, V., & Hernquist, L., 2014, MNRAS, 445, 3475

[192] Muñoz, D. J., Lai, D., 2016, ApJ, 827, 43

[193] Murray, C. D., Dermott, S. F., 1999, *Solar System Dynamics*, Cambridge University Press

[194] Muzerolle, J., Furlan, E., Flaherty, K., Balog, Z., Gutermuth, R., 2013, Nature, 493, 378

[195] Narayan, R., McClintock, J. E., 2012, MNRAS, 419, L69

[196] Nelson, R. P., 2003, MNRAS, 345, 233

[197] Nelson, R. P., & Papaloizou, J. C. B. 2004, MNRAS, 350, 849

[198] Neuhäuser, R., Krämer, S., Mugrauer, M., Köhler, R., Schmidt, T. O. B., Ammler-von Eiff, M., Alves, J., Fiedler, S., Vogt, N., 2009, A&A, 496, 777

[199] Nixon, C., Lubow, S. H., 2015, MNRAS, 448, 3472

[200] Okuzumi, S., & Ormel, C. W. 2013, ApJ, 771, 43

[201] O'Neill, S. M., Reynolds, C. S., Miller, M. C., 2009, ApJ, 693, 1100

[202] O'Neill, S. M., Reynolds, C. S., Miller, M. C., Sorathia, K. A., 2011, ApJ, 736, 107

[203] Oda, H., Machida, M., Nakamura, K. E., Matsumoto, R., 2010, ApJ, 712, 639

[204] Orosz, J. A., et al., 2012, Science, 337, 1511

[205] Ortega-Rodríguez, M., Silbergleit, A., Wagoner, R., 2008, GApFD, 102, 75

[206] Ortega-Rodríguez, M., Wagoner, R., 2000, ApJ, 537, 922

[207] Ostriker, E. C., & Shu, F. H. 1995, ApJ, 447, 813

[208] Owen, J. E., & Kollmeier, J. A. 2017, MNRAS, 467, 3379

[209] Paardekooper, S.-J., Lesur, G., & Papaloizou, J. C. B. 2010, ApJ, 725, 146

[210] Paardekooper, S.-J., Leinhardt, Z. M., Thébault, P., Baruteau, C., 2012, ApJL, 754, L16

[211] Paczyński, B., 1977, ApJ, 216, 822

[212] Paczyński, B., Witta, P. J., 1980, A&A, 88, 23

[213] Paczyński, B., Rudak, B., 1980, AcA, 30, 237

[214] Papaloizou, J., Pringle, J. E., 1977, MNRAS, 181, 441

[215] Papaloizou, J. C. B., Lin, D. N. C., 1988, ApJ, 331, 838

[216] Papaloizou, J. C. B., Nelson, R. P., Masset, F., 2001, A&A, 366, 263

[217] Pelupessy, F. I., Portegies Zwart, S., 2013, MNRAS, 429, 895

[218] Pérez, L. M., Isella, A., Carpenter, J. M., & Chandler, C. J. 2014, ApJL, 783, L13

[219] Petrovich, C., Rafikov, R. R., 2012, ApJ, 758, 33

[220] Pichardo, B., Sparke, L. S., Aguilar, L. A., 2005, MNRAS, 359, 521

[221] Pierens, A., Nelson, R. P., 2007, A&A, 472, 993

[222] Popham, R., & Narayan, R. 1991, ApJ, 370, 604

[223] Pringle, J. E., 1981, ARA&A, 19, 137

[224] Raettig, N., Klahr, H., & Lyra, W. 2015, ApJ, 804, 35

[225] Rafikov, R. R. 2002, ApJ, 572, 566

[226] Rafikov, R. R., 2013, ApJL, 764, L16

[227] Rafikov, R. R., 2015, ApJ, 804, 62

[228] Rafikov, R. R. 2016, ApJ, 830, 7

[229] Ragusa, E., Lodato, G., Price, D. J., 2016, MNRAS, 460, 1243

[230] Ragusa, E., Dipierro, G., Lodato, G., Laibe, G., & Price, D. J. 2017, MNRAS, 464, 1449

[231] Regály, Z., Juhász, A., Sándor, Z., & Dullemond, C. P. 2012, MNRAS, 419, 1701

[232] Rein, H., Latter, H. N., 2013, MNRAS, 431, 145

[233] Remillard, R. A., McClintock, J. E., 2006, ARA&A, 44, 49

[234] Reynolds, C. S., Miller, M. C., 2009, ApJ, 692, 869

[235] Rezzolla, L., Yoshida, S., Maccarone, T. J., Zanotti, O., 2003, MNRAS, 344, L37

[236] Roccatagliata, V., Ratzka, T., Henning, T., Wolf, S., Leinert, C., Bouwman, J., 2011, A&A, 534, 33

[237] Roedig, C., Sesana, A., Dotti, M., Cuadra, J., Amaro-Seoane, P., Haardt, F., 2012, A&A, 545, A127

[238] Romanova, M. M., & Owocki, S. P. 2015, Space Sci. Rev., 191, 339

[239] Rothstein, D. M., Lovelace, R. V. E., 2008, ApJ, 677, 1221

[240] Russo, M., Thompson, C., 2015, ApJ, 813, 81

[241] Sano, T., Miyama, S. M., Umebayashi, T., Nakano, T., 2000, ApJ, 543, 486

[242] Schmidt, J., Salo, H., Spahn, F., Petzschmann, O., 2001, Icarus, 153, 316

[243] Schmit, U., Tscharnuter, W. M., 1995, Icarus, 115, 304

[244] Schnittman, J. D., Bertschinger, E., 2004, ApJ, 606, 1098

[245] Sesana, A., Vecchio, A., Colacino, C. N., 2008, MNRAS, 390, 192

[246] Shakura, N. I., Sunyaev, R. A., 1973, A&A, 24, 337

[247] Shcherbakov, R. V., McKinney, J. C., 2013, ApJL, 774, L22

[248] Shi, J.-M., Krolik, J. H., Lubow, S. H., Hawley, J. F., 2012, ApJ, 749, 118

[249] Shi, J.-M., & Krolik, J. H., 2015, ApJ, 807, 131

[250] Silsbee, K., Rafikov, R. R., 2015, ApJ, 808, 58

[251] Simon, M., Dutrey, A., Guilloteau, S., 2000, ApJ, 545, 1034

[252] Simon, J. B., Lesur, G., Kunz, M. W., & Armitage, P. J. 2015, MNRAS, 454, 1117

[253] Springel, V., 2010, MNRAS, 401, 791

[254] Stapelfeldt, K. R., Krist, J. E., Ménard, F., Bouvier, J., Padgett, D. L., Burrows, C. J., 1998, ApJ, 502, L65

[255] Stapelfeldt, K. R., Ménard, F., Watson, A. M., Krist, J. E., Dougados, C., Padgett, D. L., Brandner, W., 2003, ApJ, 589, 410

[256] Stella, L., Vietri, M., Morsink, S. M., 1999, ApJ, 524, L63

[257] Surville, C., Mayer, L., & Lin, D. N. C. 2016, ApJ, 831, 82

[258] Tagger, M., Varnière, P., 2006, ApJ, 652, 1457

[259] Takeuchi, T., Miyama, S. M., Lin, D. N. C., 1996, ApJ, 460, 832

[260] Tanga, P., Babiano, A., Dubrulle, B., Provenzale, A., 1996, Icarus, 121, 158

[261] Tsang, D., Lai, D., 2008, MNRAS, 387, 446

[262] Tsang, D., Lai, D., 2009a, MNRAS, 396, 589

[263] Tsang, D., Lai, D., 2009b, MNRAS, 400, 470

[264] Tsang, D. 2011, ApJ, 741, 109

[265] Turner, N. J., Fromang, S., Gammie, C., Klahr, H., Lesur, G., Wardle, M., Bai, X.-N., 2014, Protostars and Planets VI, University of Arizona Press

[266] Urpin, V., Brandenburg, A., 1998, MNRAS, 294, 399

[267] van der Marel, N., van Dishoeck, E. F., Bruderer, S., et al. 2013, Science, 340, 1199

[268] van der Marel, N., Cazzoletti, P., Pinilla, P., & Garufi, A. 2016, ApJ, 832, 178

[269] Varnière, P., & Tagger, M. 2006, A&A, 446, L13

[270] Ward, W. R. 1986, Icarus, 67, 164

[271] Ward, W. R. 1997, Icarus, 126, 261

[272] Ward, W. R., & Hourigan, K. 1989, ApJ, 347, 490

[273] Wellons, S., Zhu, Y., Psaltis, D., Narayan, R., McClintock, J. E., 2014, ApJ, 785, 142

[274] Welsh, W. F., Orosz, J. A., Carter, J. A., et al. 2012, Nature, 481, 475

[275] Welsh, W. F., Orosz, J. A., Short, D. R., et al. 2015, ApJ, 809, 26

[276] Winn, J. N., Holman, M. J., Johnson, J. A., Stanek, K. Z., Garnavich, P. M., 2004, ApJ, 603, L45

[277] Wu, Y., Goldreich, P., 2002, ApJ, 564, 1024

[278] Wyithe, J. S. B., Loeb, A. 2003, ApJ, 590, 691

[279] Yuan, F., Lin, J., Wu, K., Ho, L. C. 2009, MNRAS, 395, 2183

[280] Zhu, Z., Hartmann, L., Gammie, C. F., 2010a, ApJ, 713, 1143

[281] Zhu, Z., Hartmann, L., Gammie, C. F., Book, L.G., Simon, J.B., Engelhard, E., 2010b, ApJ, 713, 1134

[282] Zhu, Z., & Baruteau, C. 2016, MNRAS, 458, 3918

[283] Zhu, Z., & Stone, J. M. 2014, ApJ, 795, 53